Design of Experiments and Reliability Growth on Repairable Systems

by

Dustin Taylor

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2023 by the
Graduate Supervisory Committee:

Douglas Montgomery, Co-Chair
Rong Pan, Co-Chair
Steve Rigdon
Laura Freeman
Ashif Iquebal

ARIZONA STATE UNIVERSITY

August 2023

ABSTRACT

Reliability growth is not a new topic in either engineering or statistics and has been a major focus for the past few decades. The increasing level of high-tech complex systems and interconnected components and systems implies that reliability problems will continue to exist and may require more complex solutions. The most heavily used experimental designs in assessing and predicting a systems reliability are the "classical designs", such as full factorial designs, fractional factorial designs, and Latin square designs. They are so heavily used because they are optimal in their own right and have served superbly well in providing efficient insight into the underlying structure of industrial processes. However, cases do arise when the classical designs do not cover a particular practical situation. Repairable systems are such a case in that they usually have limitations on the maximum number of runs or too many varying levels for factors. This research explores the D-optimal design criteria as it applies to the Poisson Regression model on repairable systems, with a number of independent variables and under varying assumptions, to include the total time tested at a specific design point with fixed parameters, the use of a Bayesian approach with unknown parameters, and how the design region affects the optimal design. In applying experimental design to these complex repairable systems, one may discover interactions between stressors and provide better failure data. Our novel approach of accounting for time and the design space in the early stages of testing of repairable systems should, theoretically, in the final engineering design improve the system's reliability, maintainability and availability.

DEDICATION

To my wife Kaitlin, my big sis Jamie, and my first born Caroline

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my committee who have generously given their time and expertise to better my work. A special thank you goes to my co-chairs, Dr. Pan and Dr. Montgomery, for their countless hours of reflecting, reading, and encouraging. They have been instrumental in my professional development. I have enjoyed the friendly chats at the end of our meetings, the get-togethers, and the swim/BBQ. I am also grateful to Dr. Rigdon for his personal and professional guidance, and the free insights into the world of hockey. I would be remiss in not mentioning my family, whose belief in me has kept my spirits and motivation high during this process. To my wife, Kaitlin, your love and understanding helped me through the dark times. Without you believing in me, I never would have made it. To my big sister, Jamie, for whom I would have never had this adventure and lastly my two dogs, Whiskey and Rye, for all their entertainment and emotional support!

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

1.1 Background and Motivation

Reliability is one of the key performance parameters that is tracked during the U.S Department of Defense (DoD) acquisition process. It is the innate capability of a system to perform the intended function(s) over a period of time. In early stages of development, systems typically will fail to meet the customers reliability requirements. To improve reliability, and to create *reliability growth*, a series of test events are used to identify failure modes. These failure modes are analyzed and fixed and testing continues. This process is referred to as "Test-Analyze-Fix-Test" or TAFT.

There are two categories of systems for reliability models. The first is *non-repairable* where a system is discarded after failure, such as light bulbs, batteries, and even televisions. These are considered "one-shot" devices with only a single failure being observed. The second are *repairable* systems where the system can be restored to an operating condition without replacing the entire system. Failures to the system are repaired by fixing or replacing components, such as replacing a starter on a car or replacing a cooling fan on your computer. Multiple failures will be observed for each system during testing. Our research starts from experimental designs for non-repairable systems and is extended to *repairable* systems.

The expected lifetime and mean time between failure (MTBF) helps address the reliability area of concern. It is well known that reliability estimates using data from developmental testing (DT) are often optimistic estimates of reliability under

operationally relevant conditions. One problem in using this optimistic estimate, is the systems will be viewed ready for operational testing when, in some cases, their operationally relevant reliability may be substantially smaller. As a result, either testers will need to find a large number of design defects during operational testing, a goal inconsistent with the structure of operational testing, or a system with deficient reliability will be promoted to full-rate production (Council 2014). Design changes at the operational testing phase is rarely an option due to the inefficiency and cost associated with such decision. In other words, fixing a problem at the later stage of system design is much more expensive than at an earlier stage.

To mature the reliability of a complex system under development, we must detail a reliability growth plan that depicts how the systems reliability is expected to increase over the testing period (Ellner and Hall 2006). Current methods involve selecting a model based on the type of data and the method of analysis. A Poisson Regression model is one such method. Poisson regression models have been around for a while and are fairly easy to use yet has flexibility. These models can represent either homogeneous or non-homogeneous process. Yet what these current methods lack, and this research investigates, is combining the model, design space, and testing time through a Bayesian approach in determining an optimal test design.

Furthermore, advancements in science, technology, and engineering has led to manufactures developing highly sophisticated and very reliable products with long life time, to include things like computer equipment, bombs, and silicone rubber seals. The reliability of a product determines the quality and competence of the product and therefore, the manufacturers put in millions for assessment of a products reliability, formulating new designs and so on of their products (Nassar, Dey, and Nadarajah 2021). One drawback is that for highly reliable products, which have a longer life

time, it is often not possible to conduct life testing experiments at the normal design, or use condition, to obtain failure data. To overcome this issue, we turn to accelerated life testing (ALT) in obtaining the necessary failure data. During ALT, the systems are exposed to higher levels of stress factors such as temperature, force, or humidity, so that we may observe more failures. It is usually more expensive to conduct tests in an accelerated manner, and one must account for the cost and time available to test at each design point.

The question now remains how to ensure the early stages of testing invoke the necessary failures, as a system matures, in order to properly determine the reliability of the system. Härtler (1989) states "...even when the model fits the data in a perfect way and when sufficient statistical methods are available any analysis remains incomplete and any conclusions remain questionable unless the quantitative analysis is not completed by investigation of failure causes." It is not enough to know the cause of failure, but to ensure we've tested the system in an array of possible failure causes. How do we know we uncovered all of, or enough failures, to accurately predict the systems reliability? To accomplish this, this research uses D-Optimality designs and a Bayesian methodological framework. This research takes a novel approach in assuming the failure rate is a function of the design space, and then investigates how the design space, that is the design region, is affected. This research further embeds and accounts for the testing duration, time, at each specific design point of the design space.

1.2 Poisson Process and Models

1.2.1 Homogeneous Poisson Process

Aminzadeh (2013) states that the Homogeneous Poisson Process (HPP) is a special case of a renewal process which is often used in practice. It is one of the simplest point process models when modeling the rate of occurrence (repair rate) of failures for a repairable system and counts the number of failures during a specific time period. This model comes about when the interarrival times between failures are iid according to the exponential distribution, with parameter (intensity) $\lambda > 0$. That is, a counting process $N(t), t \in [0, \inf)$ is a Homogeneous Poisson Process if the rate (or intensity) λ is a constant.

Despite the simplicity of this model, it is widely used for repairable equipment and systems throughout industry. Justification for this comes, in part, from the shape of the empirical Bathtub Curve. Most systems (from complex tools, vehicles, and equipment) spend most of their "lifetimes" operating in the long flat constant repair rate portion of the Bathtub Curve. The HPP is the only model that applies to that portion of the curve, so it is the most popular model for system reliability evaluation and reliability test planning (Guthrie 2020).

1.2.2 Non-Homogeneous Poisson Process

There are times when a constant arrival rate is not always ideal, specifically when developing a new system which is at the beginning portion of the Bathtub curve. In this case, we have a non-homogenous Poisson process (NHPP) where the arrival

rates are a function of and vary with time. For example, if we assume that when a component fails and the system repaired returns to the same state it was in just before the failure, we have an NHPP. A popular model for the NHPP case is the power model where the intensity becomes a function of time.

1.2.3 Crow-AMSAA Model (Power-Law Model)

In 1974, Crow Crow 1975 showed that Duane's empirical model is essentially a Nonhomogeneous Poisson Process (NHPP) with a Weibull intensity function. Even though Crow's growth parameter estimate is still interpreted as the estimate of the negative slope of a straight line on a log-log plot, the estimates of λ and β differ from Duane's procedures in that the estimation procedure is Maximum Likelihood Estimate (MLE), not least squares, thus each model's parameters correspond to different straight lines (AMSAA 2011). Though Rigdon (2002) shows that this is a common misconception and the Duane plot is not synonymous with the power-law process.

Later, Crow (1982) obtained a functional form for the expected failure intensity as a function of test time and planning parameters as they related to the expected number of failure modes and total test duration. Furthermore, the MIL-HDBK-189 approach utilizes average MTBF values over multiple test phases to depict the planned reliability growth path. This yields the Crow-AMSAA Model, also known as the power-law process and alternately referred to as the Reliability Growth Tracking Model Continuous (RGTMC) model. Under this Crow-AMSAA model, the failure intensity function is given by $\rho(t) = \lambda\beta t^{\beta-1}$ from the Weibull intensity function with λ, the scale parameter, and β, the shape parameter, being estimated through the MLE

process. Some believe β is the most important parameter since it determines whether the reliability is growing or not, but it is also the most difficult one to accurately estimate. The Crow-AMSAA model looks at the reliability growth within a particular interval of time. Assume that a particular test begins at time $t = 0$ with each failure being the end of one interval and start of another. We will then have a testing period of $0 < t_1 < t_2 < ... < t_n$. The failure intensity, λ_i, is assumed constant within each interval $[t_{i-1}, t_i]$ when a failure occurs at t_i. Therefore, the number of total failures, N_i, during the i^{th} time period has a Poisson distribution with mean $\lambda_i(t_i - t_{i-1})$. Overall, it is a piece-wise constant failure rate function and is an approximation to a continuous function.

$$Pr(N_i = n) = \frac{[\lambda_i(t_i - t_{i-1})]^n \mathrm{e}^{-\lambda_i(t_i - t_{i-1})}}{n!} \ , \quad n = 0, 1, 2, ...$$

The constant failure intensity, λ_i, assumes that the times between these failures follow an exponential distribution.

$$F(x) = 1 - \mathrm{e}^{\lambda_i x} \ , \quad x > 0$$

Now, let $N(t)$ be the cumulative number of failures over the total test time T that follows a NHPP process. Then the mean value function is

$$\theta(T) = \int_0^T \rho(y)dy \tag{1.1}$$

where

$$\rho(y) = \lambda$$

$$y \in [t_{i-1}, t_i]$$

Hence for total time T

$$Pr[N(T) = n] = \frac{[\theta(T)]^n \mathrm{e}^{-\theta(T)}}{n!} \ , \quad n = 0, 1, 2, ...$$

Now the Crow-AMSAA model assumes that $\rho(T)$ can be approximated by the Weibull failure rate function, or more precisely the Weibull distribution governs the first system failure and the Power Law model governs each succeeding system failure. From equation 1.1, the intensity and average number of failures become

$$\rho(T) = \lambda\beta T^{\beta-1}$$
$$\theta(T) = \int_0^T \rho(t)dt = \lambda T^{\beta}$$

Furthermore, the Crow-AMSAA model generalizes the no growth case of an HPP to allow for reliability growth due to corrective actions. This leads to the probability that the number of failures, $N(t)$, is given by:

$$P[N(T) = n] = \frac{(\lambda T^{\beta})^n \mathrm{e}^{-\lambda T^{\beta}}}{n!}$$

The parameter estimations for the failure data within the Crow-AMSAA model are estimated using MLE. The cumulative density function (CDF) of the i^{th} failure at t_i is:

$$F(t_i|t_{i-1}) = 1 - \mathrm{e}^{-\lambda(t_i^{\beta} - t_{i-1}^{\beta})}$$

Next we get the probability density function (pdf) of the i^{th} failure at t_i is:

$$f(t_i|t_{i-1}) = \lambda\beta(t_i^{\beta-1})\mathrm{e}^{-\lambda(t_i^{\beta} - t_{i-1}^{\beta})} \tag{1.2}$$

This, through the likelihood function,

$$L = \lambda^n \beta^n \mathrm{e}^{\lambda T^{\beta}} \prod_{i=1}^{n} t_i^{\beta-1}$$

where total time T is the last observed failure time, we then take the natural log

$$\ln L = n\ln(\lambda) + n\ln(\beta) - \lambda T^{\beta} + (\beta - 1)\sum_{i=1}^{n} \ln(t_i) \tag{1.3}$$

next taking partial derivative yield parameter estimates of $\hat{\lambda}$ and $\hat{\beta}$, given a termination time (or Total Time), T, we have:

$$\hat{\lambda} = \frac{n}{T^{\beta}} \tag{1.4}$$

$$\hat{\beta} = \frac{n}{n\ln(T) - \sum_{i=1}^{n}\ln(t_i)} \tag{1.5}$$

Furthermore, this yields a failure intensity function estimation as:

$$\hat{\rho}(T) = \hat{\lambda}\hat{\beta}T^{\hat{\beta}-1} \;=\; \hat{\beta}(\frac{n}{T})$$

Further derivation of the likelihood function can be seen in Cahoon, Sanborn, and Wilson 2021

1.2.4 AMSAA Maturity Projection Model (AMPM)

"The power law approach uses an assumed cumulative relationship between the expected number of discovered failures and test duration. The AMPM approach uses an assumed cumulative relationship between the expected number of discovered B-modes and the test duration, which gives rise to a reliability growth relationship between the expected system failure intensity and the cumulative test duration. Failure modes fall into two categories: A-modes and B-modes" (AMSAA 2011). A-modes will not have corrective action taken, while B-modes will be addressed via corrective action.

The main purpose of the AMPM approach is to provide an estimate of the projected reliability following implementation of both delayed and non-delayed fixes. This approach is based on viewing B-failure mode rates $(\lambda_1, ..., \lambda_k)$ as a random sample, call them $\wedge = (\wedge_1, ..., \wedge_K)$ from a gamma distribution, $\Gamma(\alpha, \beta)$. The failure mode's first occurrence failure times are conditioned on the λ_i's and are assumed to be

independent and exponentially distributed with means λ_i. This allows one to utilize all the B-mode times to the first observed failures during the initial Test Phase I in estimating the gamma parameters.

There are several ways in which the AMPM approach can be used to make reliability projections. First is where all fixes to B-mode failures are implemented at the end of the current phase, prior to commencing a follow-on test. The second situation is where the system under test has been maturing over the Phase due to implementing fixes during and withing the testing phase. The third situation involves projecting the system failure's intensity at a future program milestone. Further estimations and derivations can be found in the MIL-HDBK 189C (AMSAA 2011) section 6.3.

1.2.5 Other Models

Other improvements for the Crow-AMSAA model have been developed in the past decade. Gaver and Jacobs (2014) extends the NHPP model to include fault mode (FM) type differences plus the realistic probability that actuated FMs are not removed until a removal is successful. Awad (2016) proposes a method by assuming a dynamic failure intensity and modeled using a Weibull distribution. Peng et al. (2017) proposes a novel method to evaluate and predict the dynamic reliability of a repairable system subject to interval-censored problem. Kunsong and Yunxia (2021) proposed two new models, borrowing from the idea of accelerated failure-time models, that consider time-varying failure intensity in each test phase . There are many more models, such as accounting for heterogeneity in the power law NHPP (Asfaw and Lindqvist 2015) and the Modulated Power Law Process (Tang et al. 2017, and Muralidharan 2002), that the reader may find of interest but will not be discussed here.

It is important to reiterate that the current models and methods discussed above fail to account for the design of experiments. They do not account for the design space in which the failure data was derived, and the models do not account for testing time at the design points (the focal point of this dissertation).

1.3 Optimality of a Design

The optimal design of experiment is a method used to determine the best way to collect data for a particular study. It involves selecting the most appropriate sample size, sampling method, and experimental design to ensure that the results obtained are accurate and reliable.

The most heavily used designs in industry are the "classical designs" (full factorial designs, fractional factorial designs, Latin square designs, Box-Behnken designs, etc.) and have a long history of success within a variety of industries. They are so heavily used because they are optimal in their own right and have served superbly well in providing efficient insight into the underlying structure of industrial processes (Guthrie 2020). Cases do arise when the classical designs do not cover a particular practical situation. Such cases, listed by the National Institute of Standards and Technology (Guthrie 2020) include limitations on the maximum number of runs, too many varying levels for factors, or the assumption of an underlying complicated model; this is where optimal designs come in.

D-optimal designs are one form of optimal designs and a focus in this paper. Montgomery (2019) states that a design that minimizes the variance of the model regression coefficients is called a D-optimal design. The D term is used because it maximizes the determinant of the Fisher information matrix of all design points within

a certain design space. This optimality criterion results in minimizing the generalized variance of the parameter estimates for a pre-specified model, such as the Poisson Regression model. As a result, the 'optimality' of a given D-optimal design is model dependent and the experimenter must specify the models' coefficients before a design can be generated. This leads us to a Bayisean approach.

1.4 Bayesian Approach

The assumption of normality is usually tied to the design and analysis of an experimental study. However, when dealing with lifetime testing and censoring at fixed time intervals, we can no longer assume the outcomes will be normally distributed. This generally requires the use of optimal design techniques to construct the test plan for specific distribution of interest. Optimal designs in this situation depend on the parameters of the distribution which are generally unknown a priori. The Bayesian approach can be applied to this situation where the optimal design depends on unknown parameters. The Bayesian approach involves selecting a prior distribution to reflect our knowledge about parameters before data are collected, and then after data are collected, applying Bayes theorem to give the updated probability distribution of the parameters given the data. It forces us to quantify our belief about parameters in terms of their uncertainties. This, along with the model and number of predictors, can be used to derive the D-optimal design for an allowed number of experimental runs.

This approach to experimental design offers many advantages over other approaches, the most notable of which is the ability to optimize design criteria that are functions of the posterior distribution and that can be easily tailored to the experimenter's

design objective (Ryan et al. 2016). Tian et al. (2022) state that when facing reliability data with limited information (e.g., a small number of failures), there are strong motivations for using Bayesian inference methods. Additional details in the construction of Bayesian models have been explored in Burghaus and Dette (2014), Dette (2007), Dette and Neugebauer (1997), and Chaloner and Larntz (1992).

1.5 Structure of the Dissertation

This dissertation is organized as follows. Ch.2 explores using a Bayesian approach on various lifetime regression models to select appropriate D-optimal designs in regular and irregular design regions up-to a second order, 2 parameter model. Here we address how the levels for the predictor variables should be selected so as to maximize the information that we can gain with different levels of censoring. Ch.3 explores the optimal designs for repairable systems that follow the HPP model over discrete design space. We explore expanding the design space as well as accounting for testing time at each design point. We then briefly touch on the Non-homogeneous case in Ch.4 and a short justification to optimal designs in Ch.5 before concluding our research in Ch.6. Each chapter will include separate introductions with their own, applicable, literature review.

Chapter 2

BAYESIAN D-OPTIMAL DESIGN FOR LIFE TESTING WITH CENSORING

2.1 Introduction and Literature Review

Suppose we are planning an experiment where the outcome is a lifetime and is subject to censoring at a fixed time. The settings of the predictor variable are constrained to be in some region R of the design space. Given a fixed sample size n (which is equivalent to the number of experimental runs), how should the levels of the predictor variables be selected in each experimental run so as to provide the greatest information from the experiment? This problem of optimal experimental design has been studied thoroughly when the outcome can be assumed to be normally distributed (see, e.g., Montgomery 2019; Ryan 2013; Cobb 1998; Rasch et al. 2011; Shah and Sinha 2012). But, for lifetimes, the normal distribution is usually unreasonable; in addition, life testing experiments must deal with censoring, that is, the termination of testing before all units have failed.

For the normal theory case, the situation is simplified in that the optimal design is often dependent on the model, but not on the parameters in the model. For example, an optimal design for a first-order model in two variables is independent of the intercept β_0 and the slope parameters, β_1 and β_2. It is also independent of the error variance σ^2. When we step outside of the normal-linear model, the situation becomes more complicated. For instance, assuming that the lifetime has a Weibull distribution (a useful and widely applicable reliability model for life test planning), the optimal design depends on the value of the shape parameter. This creates a catch:

an experimenter is running the experiment so as to learn about the parameters of the assumed distribution, but only to realize that, to obtain the desirable optimal experimental design, the values of these parameters must be known beforehand.

Roy and Pradhan (2017) considers designs for life-time plans by developing a C-optimal Bayesian censoring scheme focusing on the estimation precision of a particular lifetime quantile. They propose a generic algorithm to obtain the optimal censoring schemes under two different scenarios for the Weibull and log-normal models but do not address predictor variables as we do in our paper. Zohbi, Wainakh, and Arafeh (2016) develop an algorithm for finding D-optimal designs when the response is a lifetime following a Weibull distribution. While censoring is taken into account, they assume a fixed and known shape parameter for the Weibull distribution, which is a strong assumption. Our approach described here assumes that the shape parameter κ is unknown and must be estimated from the data.

The Bayesian approach can be applied to this situation where the optimal design depends on unknown parameters. The Bayesian approach involves selecting a prior distribution to reflect our knowledge about parameters before data are collected, and then after data are collected, applying Bayes theorem to give the updated probability distribution of the parameters given the data. It forces us to quantify our belief about parameters in terms of their uncertainties. This knowledge can then be used to select a robust experimental design. Additional details in the construction of Bayesian models have been explored in Burghaus and Dette (2014), Dette (2007), Dette and Neugebauer (1997), and Chaloner and Larntz (1992).

Suppose that the lifetime of the i-th run is T_i and the vector of predictor variables is $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]^t$. We assume a model of the form

$$T_i|\boldsymbol{x}_i \sim \text{WEIBULL}(\theta = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \kappa))$$

This is equivalent to assuming that the logarithm of lifetime has a smallest extreme value (SEV) distribution with location parameter μ and scale parameter σ. If we let $Y_i = \log T_i$, then we are effectively assuming

$$Y_i|\boldsymbol{x_i} \sim \text{SEV}(\mu = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \sigma = 1/\kappa). \tag{2.1}$$

The method for finding D-optimal designs that we propose in this article can be applied to models of any order. We illustrate the method in Section 2.5, for situations of up to a two predictors and a full second-order model. The most general model that we illustrate is

$$Y_i|\boldsymbol{x_i} \sim \text{SEV}(\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2, \sigma = 1/\kappa). \tag{2.2}$$

Thus, the parameters of the SEV are related to the parameters of the Weibull as follows:

$$\begin{aligned} \mu &= \log\theta = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} \\ \sigma &= 1/\kappa. \end{aligned}$$

Note that if the predictors $x_1, \cdots, x_p$ are environmental stress variables, which accelerate a test unit's failure process, then such reliability tests are called accelerated life tests (ALTs). In ALT, the value ranges of parameters $\beta_1, \cdots, \beta_p$ are typically restrictive, because higher stress levels are supposed to shorten a test unit's lifetime. In this paper, however, we do not limit our models to ALT models. These predictors, for example, can be product design variables or manufacturing process variables, thus their effect coefficients would not be constrained to a certain range.

As with most life testing experiments, we assume that the test can be terminated before all units have failed. Here we assume that all survival observations are censored

at the same time t_c. On the log scale, this is equivalent to censoring at time $y_c = \log t_c$. Denote δ as an indicator variable for censoring,

$$\delta_i = \begin{cases} 1, & \text{observation } i \text{ is a failure} \\ 0, & \text{observation } i \text{ is censored} \end{cases}$$

The problem to be addressed is how the levels for the predictor variables should be selected so as to maximize the information that we can gain from the experiment. If the Weibull distribution is assumed for lifetimes, or equivalently, the SEV distribution for log lifetimes, then the optimal design depends on the values of the parameters in these distribution models. As we mentioned previously, the normal-linear model avoids this dilemma, but for design a life testing, this problem is unavoidable.

There have been different approaches taken to circumvent this dilemma. Park and Yum (1996) determined various pre-estimates for these parameters and then found a separate test design for each individual pre-estimate. This produces different designs based on predetermined values. Others have focused on using the log normal distribution and the distribution transformation methods that prevent the parameter-dependent problem when determining the test points (Zhang, Yang, and Kong 2021). Also, some researchers used predetermined stress levels based on single-stress tests, previous tests of similar units, prior knowledge on life–stress relationship, or equipment capacities (Nasir and Pan 2017; Pan, Yang, and Seo 2015; Zhu and Elsayed 2013; Monroe et al. 2010; Monroe and Pan 2008).

Our way forward is to take the Bayesian approach and express our beliefs about parameters in their prior distributions. If we assume an SEV distribution for log lifetimes as in Eq. (2.1), then the parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$. Let $p(\boldsymbol{\theta})$ denote the prior distribution for $\boldsymbol{\theta}$ used for the purpose of designing the experiment. Hong

et al. 2015 suggest that it may be reasonable to have separate priors for the design stage and the analysis stage. For now, $p(\boldsymbol{\theta})$ represents the prior at the design stage.

Suppose ξ is a design of the form

$$\xi = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

where x_{ij} is the i-th level for variable j. Note that this is different from the model matrix, which depends on the assumed model. For example, if a second-order model with intercept is assumed, then the first column of the design matrix would consist entirely of ones, and there would also be columns for each x_j^2 and for all combinations of $x_{j_1}x_{j_2}$. The model matrix can always be constructed from the design matrix ξ by augmenting the matrix with the appropriate columns given ξ and the model. The point is that, to find an optimal design, we need to only determine the best ξ, which embodies the settings of all predictors. Once this is determined, the usual model matrix can be obtained. Additional information on Bayesian designs can be found in Chaloner and Verdinelli 1995 and DasGupta 1996.

Let $c(t, \xi)$ denote some design criterion for the case when we observe data (t) obtained from applying design ξ. In our case, we take $c(t, \xi)$ to be the determinant of the estimated covariance matrix for the estimate of $\boldsymbol{\theta}$. We can use the preposterior

expectation of $c(t, \xi)$:

$$\begin{aligned} C(\xi) &= E_{t|\xi}\big(c(t,\xi)\big) && (2.3)\\ &= \int_{\boldsymbol{\Theta}} c(\hat{\boldsymbol{\theta}},\xi)\ p(\hat{\boldsymbol{\theta}})\ d\hat{\boldsymbol{\theta}} && (2.4)\\ &\approx \int_{\boldsymbol{\Theta}} c(\boldsymbol{\theta},\xi)\ p(\boldsymbol{\theta})\ d\boldsymbol{\theta}. && (2.5) \end{aligned}$$

In going from equation (2.3) to (2.4) we have assumed that the criterion c depends on the data t only through the estimate $\hat{\boldsymbol{\theta}}$. The approximation from (2.4) to (2.5) was suggested by Hong et al. 2015; they describe the approximation as "The predictive distribution of $\hat{\boldsymbol{\theta}}$ is a convolution of the test planning prior distribution of $\boldsymbol{\theta}$ and the distribution arising from the estimation of $\boldsymbol{\theta}$ from the data and will approach the test planning prior distribution as the sample size increases." The last expression has the interpretation as the criterion function c averaged across the joint prior distribution for $\boldsymbol{\theta}$.

As mentioned above, many articles have explored the cases of normal outcomes, single censoring levels or regular shaped grid. Others have explored outcomes with a Weibull distribution with presumed lifetime model. This article expands upon these ideas by exploring and deriving optimal designs for lifetime testing when the outcomes are assumed to have a Weibull distribution, multiple censoring levels and irregularly shaped design regions. We explore the case where the priors have discrete uniform supports and apply the coordinate exchange algorithm to produce the best Bayesian D-optimal design. We go further by taking into account different censoring levels, as well as irregularly shaped design grids, and explore how these may affect the optimal design.

2.2 Optimality Criteria

Lindley (1956) suggested that an optimality criterion should involve the expected gain in the Shannon information. This is equivalent to maximizing the expected Kullback-Leibler distance between the prior and posterior distributions. Thus, the objective is to maximize

$$U(\xi) = \int_{\mathcal{Y}} \int_{\Theta} \log \frac{p(\boldsymbol{\theta}|\boldsymbol{y}, \xi)}{p(\boldsymbol{\theta})} \; p(\boldsymbol{y}, \boldsymbol{\theta}|\xi) \; \mathrm{d}\boldsymbol{\theta} \; \mathrm{d}\boldsymbol{y} \tag{2.6}$$

where $\boldsymbol{y}$ is the observed data. This is approach is explored more thoroughly in Xu and Tang 2015.

Another approach for log-location-scale families, to include the SEV, is to minimize the volume of a posterior credible region, averaged over the prior distribution. A large sample approximation to the posterior covariance matrix was given by Hong et al. (2015); they give the approximation (using our notation)

$$V(\boldsymbol{\theta}|\boldsymbol{y}, \xi) \approx \left[S^{-1} + \hat{I}_{\boldsymbol{\theta}}(\xi) \right]^{-1}$$

where S is the variance-covariance matrix of prior distribution and $\hat{I}_{\boldsymbol{\theta}}(\xi)$ is the information matrix evaluated at $\hat{\boldsymbol{\theta}}$, which is the estimator of $\boldsymbol{\theta}$.

Following the reasoning in Hong et al. 2015, we can derive an approximation to the expected value of a function h of the covariance matrix:

$$\begin{aligned} U_h(\xi) &\approx E\left(h\left(\left[S^{-1} + \hat{I}_{\boldsymbol{\theta}}(\xi) \right]^{-1} \right) \right) \\ &\approx \int h\left(\left[S^{-1} + \hat{I}_{\boldsymbol{\theta}}(\xi) \right]^{-1} \right) \; p_0(\hat{\boldsymbol{\theta}}) \; d\hat{\boldsymbol{\theta}} \\ &\approx \int h\left(\left[S^{-1} + I_{\boldsymbol{\theta}}(\xi) \right]^{-1} \right) \; p(\boldsymbol{\theta}) \; d\boldsymbol{\theta}. \end{aligned}$$

Here p_0 is the PDF (probability density function) of the estimator $\hat{\theta}$ and $p(\boldsymbol{\theta})$ is the prior PDF for $\boldsymbol{\theta}$ with $\hat{\theta}$ approaching the prior as the sample sizes increase.

If the objective is to minimize the determinant of the large-sample approximation to the posterior covariance matrix then the utility of design ξ becomes

$$U_D(\xi) = \int_\Theta \det\left(S^{-1} + \hat{I}_{\boldsymbol{\theta}}(\xi)\right)^{-1} \; p(\boldsymbol{\theta}) \; \mathrm{d}\boldsymbol{\theta} \tag{2.7}$$

We would then want a design that minimizes $U_D(\xi)$. This is similar to the approach taken by DuMouchel and Jones (1994).

If we take a prior across a discrete set of points in the parameter space, the D-criterion becomes

$$U_D(\xi) = \sum_{\beta_0}\sum_{\beta_0}\cdots\sum_{\beta_m}\sum_{\kappa} \det\left(S^{-1} + \hat{I}_{(\beta_0,\beta_1,\ldots,\beta_m,\kappa)}(\xi)\right)^{-1} \; p(\beta_0,\beta_1,\ldots,\beta_m,\kappa). \tag{2.8}$$

The computations are more stable if we use the reciprocal $1/U_D(\xi)$ as the criterion, which we then try to maximize. Although any discrete prior would work, for our examples, we take a discrete uniform prior across all values in the prior support; in effect, this takes the harmonic average of the Bayesian D criterion at each point in the prior parameter space. Additional Bayesian utility functions for parameter estimation are discussed in Ryan et al. (2016).

2.3 Selection of Prior Distributions

In order to design a life testing experiment, we must express prior distributions for the model parameters. We will put these parameters into three categories:

1. the shape parameter κ,
2. the "intercept" parameter β_0, and
3. the "slope" parameters $\beta_1, \beta_2, \ldots, \beta_m$.

The shape parameter κ is perhaps the easiest parameter for which to specify a prior. In most life testing experiments, the hazard function is increasing but it is often concave down. Since the Weibull's hazard function is

$$h(t) = \frac{\kappa}{\theta}\left(\frac{t}{\theta}\right)^{\kappa-1} \qquad t > 0,$$

the increasing/concave down presumption implies that the exponent on t should be between 0 and 1, that is

$$0 \leq \kappa - 1 \leq 1.$$

Thus,

$$1 \leq \kappa \leq 2.$$

A prior with support on $[1, 2]$ could therefore be appropriate. To be safe, the user might extend the support to an interval such as $[0.8, 2.2]$.

Next, consider the intercept parameter β_0. The mean of the Weibull distribution is

$$E(T) = \theta\Gamma\left(1 + \frac{1}{\kappa}\right).$$

The expression $\Gamma\left(1 + \frac{1}{\kappa}\right)$ is decreasing for κ belonging to the interval $[1, 2]$ having the value of 1 when $\kappa = 1$ and 0.8862 at $\kappa = 2$. (This function decreases until reaching approximately 0.8856 when $\kappa \approx 2.17$; it then increases monotonically toward 1 as $\kappa \to \infty$.) Thus, the expected lifetime is slightly below the value of θ, with the factor being no less than 0.8862 when κ is in the interval $[1, 2]$. At the center of a design, where all coded predictors are set to 0, $\theta = \exp(\beta_0)$ is a slight overestimate of the mean. The parameter θ is often called the characteristic life, and it gives the reliability (or survival probability) as

$$S(\theta) = \exp(-(\theta/\theta)^{\kappa}) = \exp(-1) \approx 0.3679.$$

In other words, approximately 36.79% of all units will survive past the characteristic life regardless of the value of κ. If we suggested a wide interval for θ, say

$$100 \leq \theta \leq 1000;$$

that is, a tenfold difference in the possible values of θ, then

$$100 \leq \exp(\beta_0) \leq 1000$$

$$4.6 \leq \beta_0 \leq 6.9.$$

The parameter β_0 is a scale parameter for the Weibull distribution, so its value doesn't affect the optimal design, but it does affect the amount of censoring, i.e., the value y_c, as described below. In the examples that follow, we will take the support for the prior for β_0 to be the interval $[-1, 1]$, which leads to the prior for θ having support $[0.3679, 2.7183]$ at the center of the design.

The slope parameters $\beta_1, \beta_2, \ldots$ are the most difficult to assess. We will consider the case where we have two predictors and a full second-order model, so that

$$\log \theta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

Suppose that as we move from the center design point $(0, 0)$ to the one of the corners, say $(1, 0)$ the mean lifetime is doubled, which implies that θ is doubled. Then at the corner point $(1, 0)$,

$$\begin{aligned} \theta &= \exp(\beta_0 + \beta_1 \times 1 + \beta_2 \times 2 + \beta_{11} \times 1 + \beta_{22} \times 0^2 + \beta_{12} \times 1 \times 0) \\ &= \exp(\beta_0 + \beta_1 + \beta_{11}) \\ &= \exp(\beta_0)\ \exp(\beta_1 + \beta_{11}). \end{aligned}$$

The expected lifetime at (1,0) is then

$$E(T|x_1 = 1, x_2 = 0) = \exp(\beta_0 + \beta_1 + \beta_{11})\ \Gamma\left(1 + \frac{1}{\kappa}\right) \tag{2.9}$$

$$= \exp(\beta_0)\ \exp(\beta_1 + \beta_{11})\ \Gamma\left(1 + \frac{1}{\kappa}\right). \tag{2.10}$$

At $(0, 0)$, the expected lifetime is

$$E(T|x_1 = 1, x_2 = 0) = \exp(\beta_0)\ \Gamma\left(1 + \frac{1}{\kappa}\right)$$

If we believe the expected lifetime increases by a factor of F as we move from $(0, 0)$ to $(1, 0)$, we must have

$$\exp(\beta_0)\ \exp(\beta_1 + \beta_{11})\Gamma\left(1 + \frac{1}{\kappa}\right) = F\ \exp(\beta_0)\ \Gamma\left(1 + \frac{1}{\kappa}\right).$$

which implies that

$$F = \exp(\beta_1 + \beta_{11}),$$

or, equivalently,

$$\beta_1 + \beta_{11} = \log F.$$

This gives us information we could use to assess our prior information about the sum $\beta_1 + \beta_{11}$, but to assess our prior information about each term individually, we must make some assumptions about the relative sizes of β_1 and β_{11}. If we assume that they are roughly equal, then

$$2\beta_1 = \log F$$

so

$$\beta_1 = \frac{1}{2}\ \log F.$$

In ALTs, a 10 fold increase in lifetime expectancy from the low to high setting of the accelerating factor is not uncommon. Taking $F = 10$ yields

$$\beta_1 = \frac{1}{2} \log 10 \approx 1.15.$$

Thus, priors for β_1 and β_{11} over the interval $[-1.15, 1.15]$ might be appropriate. Similarly, we see that priors over $[-1.15, 1.15]$ might be appropriate for β_2, β_{22}, and β_{12}. Of course, more diffuse priors can be used to achieve a more objective prior.

Life testing experiments are often done on units that are similar to previous generations of the unit, so some prior information is likely available. The prior distributions can reflect this information. When choosing a prior distribution for a parameter that is difficult to interpret directly, we can often transform the problem to one for which we can more easily interpret, select a prior for that, and then work back to the prior for the parameter. This is the trick we have employed here. The parameters β_1, β_2, ..., are difficult to interpret, so we transformed the problem to one involving the effect of a factor moving from 0 to 1. Once we assess our prior belief about this parameter, we can transform this back to obtain a prior for each of the βs.

The approach described above gives a general way of specifying reasonable upper and lower limits of parameter values for the parameters in a Weibull regression model. If no additional information is available, then we may use uniform priors for these parameters. In the Examples section that follows, we make the assumptions that the priors for the slope parameters are discrete uniform on the Cartesian product of the regions from -2 to 2. Here we have chosen priors that are somewhat more diffuse than described previously. For example, for the first-order model with one factor, we choose the prior to have equal probability assigned to each of the values in the Cartesian product given by

$$\begin{aligned} \beta_0 &\in \text{seq}(-1, 1, length.out = 5) \\ \beta_1 &\in \text{seq}(-2, 2, length.out = 5) \\ \kappa &\in \text{seq}(1, 2, length.out = 5) \end{aligned}$$

Here we are using R's notation for creating an arithmetic sequence. The priors for

the βs must reflect the knowledge about the characteristic lifetime and will depend on the time scale used to measure lifetimes.

For second-order models in two factors we use `length.out=3` in order to keep the computations manageable; this yields a discrete prior over $3^7 = 2187$ distinct grid points.

We also consider prior distributions that guarantee a monotone increasing or decreasing characteristic life. For example, if the factor(s) are life-accelerating stress factors, for which it is safe to assume higher levels of stress will lead to a shorter characteristic life. This can be done by assuming that the predictor variables are coded so that *smaller* values indicate a higher stress and slope parameters are assigned priors with lower bounds greater than 0.

Obviously, more precise prior distributions for Weibull distribution parameters can be defined when we do have some historical lifetime data or testing data from the same or similar products. As an example, consider the data from a nickel super alloy fatigue test that was analyzed in Escobar and Meeker (2006) and Rigdon et al. (2022). The lifetime of this product was measured in kilocycles and one stress variable, called pseudostress (PS), was applied during testing. It was found that a suitable life-stress model was a second-order model such as

$$\log \theta = \beta_0 + \beta_1 \log PS + \beta_2 (\log PS)^2,$$

where $\log PS$ is the logarithm of pseudostress. These regression coefficients, β_0, β_1 and β_2, were estimated to be 217.6, -85.5 and 8.48, with standard errors (se) of 62.1, 26.5 and 2.8, respectively. If we transform the log of pseudostress through

$$x = -15.6893 + 7.711884 \log(PS)$$

then the values fall between -1 and 1. Solving for $\log(PS)$, we have

$$\log PS = \frac{x + 15.6893}{7.711884}.$$

The estimated response, kilocycles to failure, is then

$$\begin{aligned}\hat{y} &= 217.6 - 85.5\left(\frac{x + 15.6893}{7.711884}\right) + 8.48\left(\frac{x + 15.6893}{7.711884}\right)^2 \\ &= 78.7541 - 6.61266x + 0.142585x^2\end{aligned}$$

where $-1 \leq x \leq 1$. Therefore, for deriving a Bayesian optimal test plan, the informative priors for these parameters could be formulated to be in the range of $\hat{\beta} \pm 2se(\hat{\beta})$, and these ranges can be further coded into [-1, 1].

2.4 Coordinate Exchange Algorithm

The coordinate exchange algorithm is described in Meyer and Nachtsheim (1995). This algorithm follows the cyclic ascent (or Gauss-Seidel) optimization method as described in Bazaraa, Sherali, and Shetty (2013). Here we summarize the method and describe how we apply this algorithm to the problem of optimal design for life tests.

Let us begin with some design ξ_0 that describes the coordinates for the predictor variables. It differs from the model matrix in that it consists of only the design points, not the columns of all ones (for the constant term) or terms for squared or cross products. The model matrix, which can be computed from the design ξ_0, is dependent on the assumed model. We consider a rectangular grid of points across the design space and focus on one of the coordinates of one design point at a time, say, the specific design choices of the first and second predictors of one design point in a two-factor model. The design criterion (e.g., the Bayesian D-criterion given in Eq. (2.7)) is then evaluated for all possible values of a single predictor in the specified grid when the

other predictor value is fixed. Then the point that yielded the largest (or smallest, depending on the criterion) is taken and assigned as the value for the corresponding row in the design ξ. This process is then repeated across all rows, yielding the design ξ_1.

The above algorithm is then repeated on the design ξ_1, yielding ξ_2. This process is continued until consecutive designs are the same. This iterative approach allows the "restarting" of the search, but retains the good features and attributes of the obtained solution (Yang et al. 2021). Often two iterations suffice, and nearly always five iterations are enough. Recall our second-order model with two factors, using `length.out=3`, yielded 2187 distinct grid points requiring evaluation.

When this algorithm is applied life testing designs, we have to consider the impact of censoring on optimal criterion evaluation. Appendix gives the details on how to derive and evaluate the expected Fisher information matrix with right censored observations. In the following examples section, multiple scenarios are considered.

2.5 Examples

In the subsections that follow, we look at the optimal designs for various situations, including one or two factors and a first-order or second-order model. We mostly consider designs on the square $[-1, 1] \times [-1, 1]$, but also touch on irregular design regions such as a trapezoidal region and a circular region. Throughout, we apply Bayesian D-optimality as the criterion and use R-Studio to run the computations. An in-depth look into this algorithm can be found in Overstall and Woods (2015) and in Overstall, Woods, and Adamou (2017).

We consider two types of priors. One is discrete uniform over $[-2, 2]$ for all of the

β_i parameters, including the intercept β_0, and discrete uniform from 1 to 2 for κ. The other is a prior that guarantees an increasing expected lifetime as a function of the factors; this prior might be of interest if the factors were accelerating factors coded so that smaller values of the factor represent higher stress.

Often, the optimal design depends on the level of censoring, particularly when the factors are accelerating factors. We have categorized the levels of censoring according to the value of y_c. Table 1 shows the values we consider and the probability of censoring at the center of the design when $\beta_0 = 0$ and $\kappa = 1.5$.

Table 1. Censoring Levels Considered in Various Examples.

y_c	P(censor)	Description
2	2×10^{-9}	Negligible
1	0.0113	Light
0	0.3699	Moderate
-1	0.8000	Heavy
-2	0.9514	Very Heavy

(a) The Probability of being censored refers to the probability that an observation at the center of the design is censored if $\beta_0 = 0$ and $\kappa = 1.5$.

2.5.1 One Predictor with First-Order Model

The simplest example to consider is that of one factor with a first-order model with negligible censoring. It might be expected that half of the runs would be allocated at the left and right endpoints of the design space, which is denoted as $[-1, 1]$. If $n = 10$, the optimal design is found to be have five points at -1 and five points at 1, which we indicate by the ordered pair $(5, 5)$; this design yields a D-utility of 1054.573. For $n = 20$ and $n = 40$, the optimal designs are $(10, 10)$ and $(20, 20)$.

Suppose, however, that the prior for β_1 is not symmetric, but rather has support only on 0 and positive values. This might be the case if the predictor variable were an accelerating variable, or for other reason we were confident that the coefficient was nonnegative. In this case, the optimal design may try to put equal numbers of runs at the endpoints, but for odd values of n the preference is to put the larger half at the low value. The reason for this seems to be that runs at the high end tend to be longer, so they are more likely to be censored, so that these runs yield less information. Under very heavy censoring, the optimal design for an even number of runs may put more points at the low end. For example, the optimal design for $n = 10$ is $(5, 5)$, but the optimal design for $n = 11$ is $(6, 5)$. The D-utility for $(6, 5)$ is 1474.605 whereas the D-utility for $(5, 6)$ is 1455.642. With very heavy censoring an even numbered sample size can yield an unbalanced design; for example with $y_c = -2$ and $n = 10$ the optimal design is $(6, 4)$ with a D-utility of 50.04983. By contrast, the design $(5, 5)$ has D utility equal to 48.00066. With $n = 80$ the optimal design is $(48, 32)$.

2.5.2 One Predictor with Second-Order Model

Consider a second-order model in a single predictor that has the feasible region of unit interval $[-1, 1]$. The budget allows for $n = 10$ runs to be selected from the interval $[0, 1]$. We apply the coordinate exchange algorithm to find the Bayesian D-optimal design. One prior support consists of all possible selections of

$$\begin{aligned} \beta_0 &\in \text{seq}(-2, 2, 0.5) \\ \beta_1 &\in \text{seq}(-2, 2, 0.5) \\ \beta_2 &\in \text{seq}(-2, 2, 0.5) \\ \kappa &\in \text{seq}(0.8, 2.2, 0.2) \end{aligned}$$

where seq(a,b,d) is the sequence from a to b in increments of d. The prior is discrete uniform over this support.

A second prior was chosen to include only those parameters which guarantee a decreasing function on the interval $[-1, 1]$. This might be reasonable when the factor is an accelerating factor for which we know that $\log \theta = \beta_0 + \beta_1 x + \beta_2 x^2$ is an increasing function from the highest stress condition ($x = -2$) to the use stress condition ($x = 1$). This is found to include those values of β_1 and β_2 that satisfy

$$-\frac{1}{2}\beta_1 \leq \beta_2 \leq \frac{1}{2}\beta_1$$

We assumed censoring at $y_c = \log t_c = 0$ or 2, corresponding to heavy and light censoring. The parameter space is the interval $[-1, 1]$. With a second-order model we would expect the optimal design to have at least three distinct points: one at the left end, one at the right and one near the middle. Without at least three points, the parameters of the model cannot be estimated. We ran three loops through the coordinate exchange algorithm with a grid of 10 intervals, or equivalently 11 design points between 0 and 1, yielding $\delta = 0.10$, and the last two designs were identical. The optimal designs are shown in Table 2 for various choices of a prior (symmetric vs. a prior that guarantees an increasing link function) and censoring (light vs. heavy).

In each case, the optimal design involves three points, $x = -1$, $x = 0$, and $x = 1$. With $n = 10$ runs, it is not possible to place the same number of points at each design point. The design is represented by an ordered triple, where the three numbers (i, j, k) indicate that i runs are to be placed at $x = -1$, j runs at $x = 0$, and k runs are placed at $x = 1$. The optimal design depended on the circumstance, with the extra point sometimes going at $x = -1$, sometimes at $x = 0$ and sometimes at $x = 1$. The utilities for various designs is shown in the last column of Table 2. Note that when the censoring is heavy, with $y_c = 0$, the utilities are much smaller than when there is

light censoring. This is due to less information contained in censored observations compared to uncensored ones.

Table 2. Optimal Designs for $n = 10$ Run Designs over Parameter Space $[-1, 1]$

Prior	**Censoring**	**Optimal Design**	**Utility**
Symmetric	Heavy ($y_c = 0$)	(4,3,3)	601.498
Symmetric	Light ($y_c = 2$)	(3,4,3)	1531.333
Guarantees increasing link	Heavy ($y_c = 0$)	(3,3,4)	464.855
Guarantees increasing link	Light ($y_c = 2$)	(3,3,4)	1439.007

The optimal design found from this algorithm is required to have runs only at the grid points. With a grid of just 11 points for each variable (121 in total), it may be that the optimal design on the continuous design space maybe slightly different from the one on the discretized grid. Since it is unlikely that moving points away from the endpoint $x = -1$ or $x = 1$ will increase a design's utility, we considered the possibility that moving the middle point to the left or right might improve the design. Table 3 shows the results of this perturbation experiment. In most cases, moving the point left or right a small amount, say 0.01, did not improve the utility of the design. For the case of a guaranteed increasing link function with heavy censoring, taking the middle point at $x = 0.01,\ 0.02,\ 0.03,\ 0.04$ yielded utilities of $465.3329,\ 465.6146,\ 465.6994,\ 465.5874$, respectively. Thus, the design with $(3, 3, 4)$ points at $x = -1, x = 0.03, x = 1$ is slightly preferred over $(3, 3, 4)$ points at $x = -1,\ x = 0,\ x = 1$ or $x = -1,\ x = 0.01,\ x = 1$.

2.5.3 Two Predictors with First-Order Model

With a prior for β_0, β_1, and β_2 symmetric about the origin the optimal design for $n = 40$ puts 10 points at each of the four corners $(-1, -1), (1, -1), (-1, 1)$, and

Table 3. Designs for Second-order Model in One Variable with a Budget of $n = 10$ Runs.

Symmetric Prior		Prior Guarantees Increasing Function	
Censoring $y_c = 0$	Censoring at $y_c = 2$	Censoring at $y_c = 0$	Censoring at $y_c = 2$
Runs at $-1, 0, 1$	Runs at $-1, 0, 1$	Runs at $-1, 0, 1$	Runs at $-1, 0, 1$
DesignUtility	DesignUtility	DesignUtility	DesignUtility
(4,3,3) 601.4985	(4,3,3) 1526.579	(4,3,3) 444.9591	(4,3,3) 1413.522
(3,4,3) 601.3223	(3,4,3) 1531.333	(3,4,3) 462.0097	(3,4,3) 1436.669
(3,3,4) 601.4985	(3,3,4) 1526.579	(3,3,4) 464.8547	(3,3,4) 1439.007
Runs at $-1, -0.01, 1$	Runs at $-1, -0.01, 1$	Runs at $-1, -0.01, 1$	Runs at $-1, -0.01, 1$
DesignUtility	DesignUtility	DesignUtility	DesignUtility
(4,3,3) 601.4153	(4,3,3) 1526.296	(4,3,3) 444.3329	(4,3,3) 1412.734
(3,4,3) 601.2124	(3,4,3) 1531.032	(3,4,3) 461.2848	(3,4,3) 1435.784
(3,3,4) 601.3609	(3,3,4) 1526.260	(3,3,4) 464.1806	(3,3,4) 1438.245
Runs at $-1, 0.01, 1$	Runs at $-1, 0.01, 1$	Runs at $-1, 0.01, 1$	Runs at $-1, 0.01, 1$
DesignUtility	DesignUtility	DesignUtility	DesignUtility
(4,3,3) 601.3609	(4,3,3) 1526.260	(4,3,3) 445.3981	(4,3,3) 1413.735
(3,4,3) 601.2124	(3,4,3) 1531.032	(3,4,3) 462.5391	(3,4,3) 1436.968
(3,3,4) 601.4153	(3,3,4) 1526.260	(3,3,4) 465.3329	(3,3,4) 1439.321

(a) All designs involve runs at three distinct points, $x = -1$, a point near $x = 0$, and $x = 1$. The design is represented by an ordered triple, where the three numbers indicate the number of runs at each point. The prior is either symmetric and flat for β_0, β_1, and β_2, or it is flat across the set of β_1 and β_2 that guarantee an increasing link function. Censoring is either heavy ($y_c = 0$) or light ($y_c = 2$).

$(1, 1)$. We will denote this design as (10,10,10,10). When the prior for β_1 and β_2 has support only on nonnegative values, then runs at the high end of either x_1 or x_2 are less attractive, because they are more likely to be censored. Depending on the amount of censoring, the optimal design may put more runs at the lower left corner ($x_1 = -1$ and $x_2 = -1$). See Table 4.

Table 4. Optimal Designs for Two Factors and a First-order Model.

		Light Censoring ($y_c = 2$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
-1	-1	11	11	12	11
1	-1	10	11	10	10
-1	1	11	10	10	10
1	1	8	8	8	9
Utility		2,367,228	2,367,228	2,359,288	2,357,229
		Heavy Censoring ($y_c = 0$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
-1	-1	11	11	12	11
1	-1	10	11	10	10
-1	1	11	10	10	10
1	1	8	8	8	9
Utility		844,958	844,958	844,308	826,128
		Heavier Censoring ($y_c = -1$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
-1	-1	14	14	13	12
1	-1	13	12	12	11
-1	1	13	12	12	11
1	1	0	2	3	6
Utility		399,423	385,753	377,213	345,831
		Very Heavy Censoring ($y_c = -2$)			
x_1	x_2	Design 1	Design 2	Design 3	Design 4
-1	-1	15	15	16	14
0.8	-1	12	13	12	13
-1	0.8	13	12	12	13
Utility		92,897	92,897	92,777	92,586

(a) The optimal designs, as determined by the coordinate exchange algorithm are in the column headed by Design 1. Designs 2, 3 and 4 are other designs that are close to the optimal design (sometimes giving exactly the same utility).

2.5.4 Two Predictors with Second-Order Model

Now consider the case of two predictor variables with a full second-order model of the form

$$T_i \sim \text{WEIBULL}(\kappa, \theta_i)$$

where

$$\theta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \beta_{11} x_{i1}^2 + \beta_{22} x_{i2}^2.$$

If the prior for $(\beta_1, \beta_2, \beta_{12}, \beta_{11}, \beta_{22})$ is symmetric about zero and the censoring

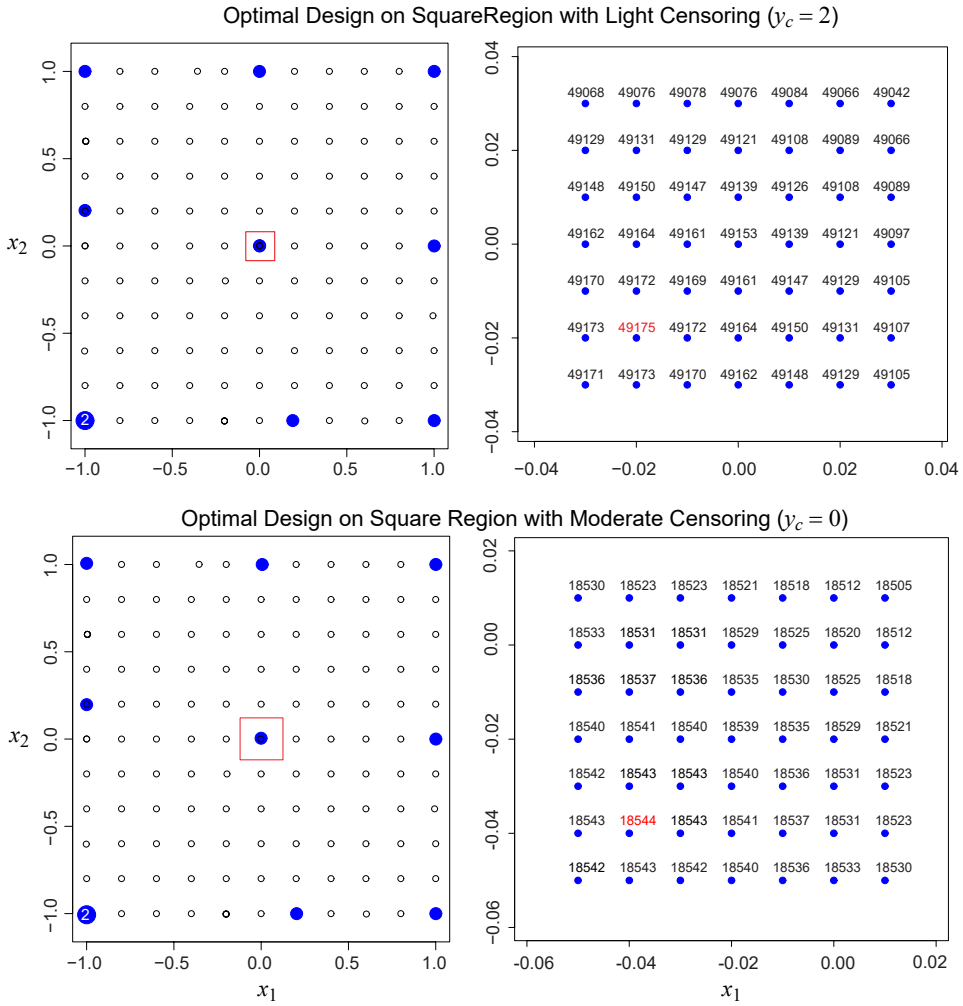
level is light, with $y_c = 2$, then the optimal design for $n = 10$ runs on a grid with 10 intervals over the domain $[-1, 1] \times [-1, 1]$ is the one shown in the top left panel of Figure 1. This design has one run at each corner, one run at the midpoint of each side, one center point run, and one additional run at one of the corners. In fact, putting two runs at any one corner and one run at every other corners yields exactly the same D criterion.

To see the effect of the fineness or coarseness of the grid, we considered moving the center point around near the middle of the design region. The "center" point was chosen to be on the grid $\{-0.03, -0.02, \ldots, 0.03\} \times \{-0.03, -0.02, \ldots, 0.03\}$. The optimal design on the grid $\{-1.0, -0.8, \ldots, 1.0\} \times \{-1.0, -0.8, \ldots, 1.00\}$ yielded a D-criterion of 49243.6. The best design obtained by moving the center point across this finer grid yielded a D-criterion of 49660.78. Thus, relative to the better design (but not necessarily the *optimal* design on an even finer grid), the optimal design obtained from the coarse grid has a relative efficiency of $49244/49661 \approx 0.99$. Thus, it seems that a grid with $n = 10$ intervals is sufficient to obtain a design that is very close to the optimum over a finer grid.

2.5.5 Two Predictors with Second-Order Model and Irregular Trapezoidal Region

Consider now a second-order model with two predictor variables over an irregular design region. Reference StatEase 2020 suggested an example which is equivalent to having a design region consisting of the unit square $[0, 1] \times [0, 1]$ subject to $x_2 \geq 1/3 - x_1/3$. This region is shown in Figure 2 with a regular grid, where the lines are parallel to the coordinate axes. The problem with this grid is that boundary points are not necessarily grid points. For example, the lower left corner point is not a grid

Figure 1. Two Factors and a Second-order Model.



(a) The prior is symmetric about the origin and the number of runs is $n = 10$. Censoring is light ($y_c = 2$) in the top two graphs and moderate ($y_c = 0$) in the bottom two. The optimal designs on the grid with 10 intervals (width of grid points is 0.2) is shown on the left. In the right two figures, we moved the single center point around to see the effect of the fineness of the grid.

point. Since optimal designs often have experimental runs at the corners, this means that the optimal design over this grid must instead put runs at the grid point nearest to the corner.

Figure 2. Optimal Design for Irregular Region with Nonparallel Grid Lines.

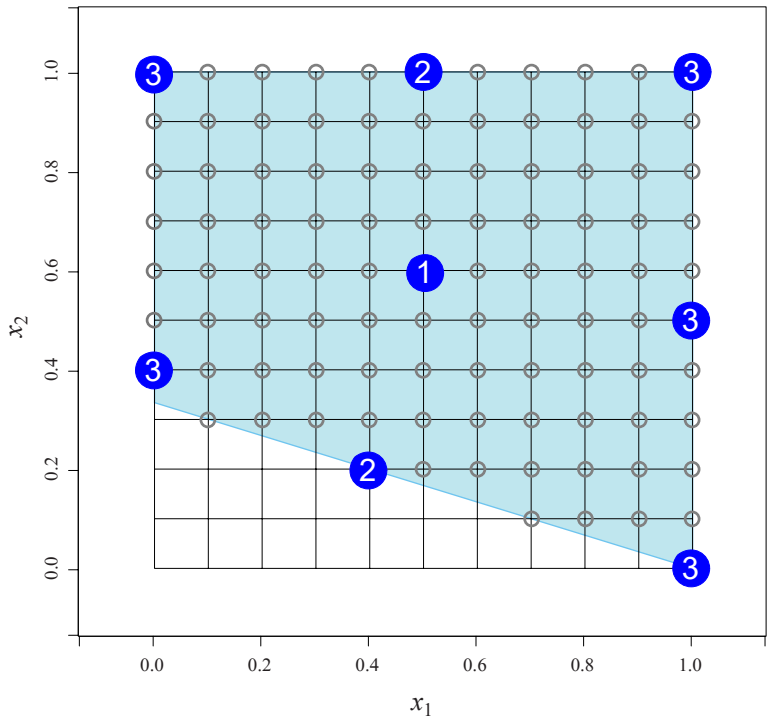
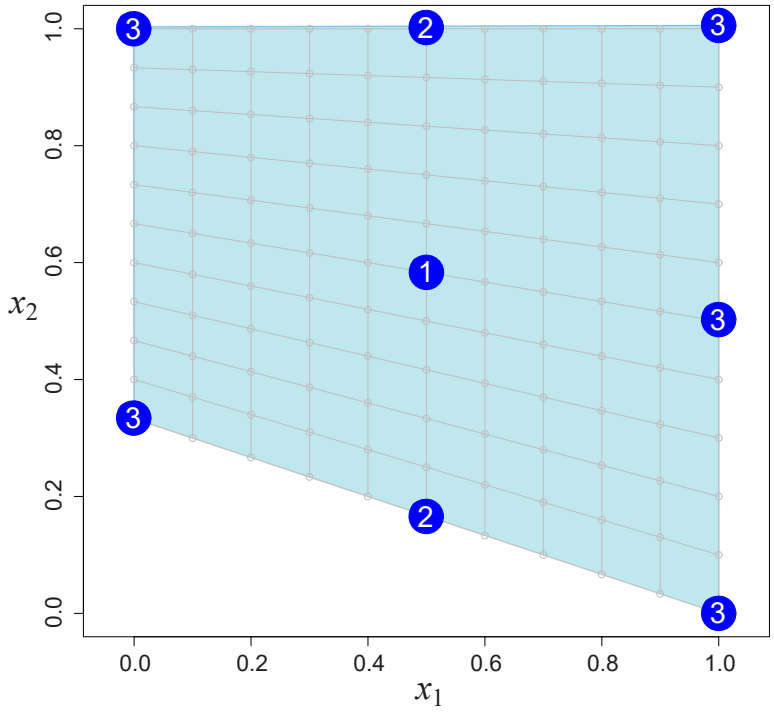


Figure 2 further shows the result of applying the coordinate exchange algorithm where possible experimental points are constrained to be on the regular grid. Runs are placed at the two nearest grid points to the lower left corner. Also the middle point on the lower boundary occurs at $x = 0.4$,not $x = 0.5$ as we might expect.

We have found that it is better to create a grid whose lines are not necessarily parallel to the coordinate axes. Basically, the idea is to divide each of the four edges into ten intervals, and then connect corresponding points at opposite ends. Such a grid is shown in Figure 3.

Figure 3. Optimal Design for Irregular Region with Grid Lines Parallel to the Coordinate Axes.



2.5.6 Two Predictors with Second-Order Model and Circular Region

Consider now a second-order model with two predictor variables over a circular design region. The mapping from a square grid to an ellipse grid can be seen below where x and y, are the coordinates within a square grid ranging [-1,1]. Additional details can be found at "https://www.xarg.org/2017/07/how-to-map-a-square-to-a-circle/," Accessed April 30, 2022.

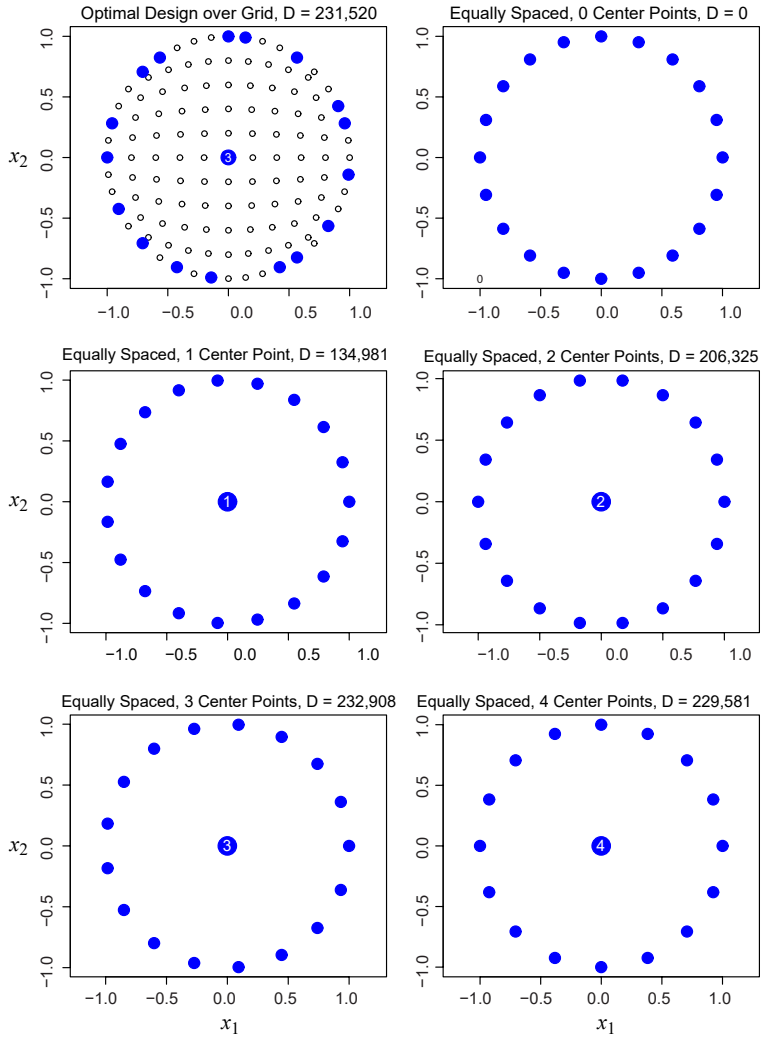
$$(x', y') = \left(x\sqrt{1 - \frac{y^2}{2}}, y\sqrt{1 - \frac{x^2}{2}} \right)$$

This circular region is shown in Figure 4 with a square grid that's been transformed, or mapped, to a circle. Because of the discreteness of the grid, it is usually impossible to obtain points that are equally spaced. For example, with a budget of 20 runs,

having 3 at the center leaves 17 to be placed on the circumference of the circle. An 11×11 grid over a square yields 40 points along the perimeter of the circle. It is not possible to place these 17 points on the perimeter in a way that they are equally spaced.

The top left graph in Figure 4 shows the grid and the optimal design over this grid. The other graphs in Figure 4 show the designs with 0, 1, 2, 3, and 4 center points, leaving 20, 19, 18, 17, and 16 points, respectively, along the perimeter. Figure 4 shows the D-criterion for each of these designs. Among these designs, the one with 3 center points has the highest D-criterion. This design is slightly better ($D = 232.908$ vs. 231.520) than the one over the grid shown in the top left figure. Figure 5 further explores the grid to circle mapping for different censoring levels; light, moderate, heavy and very heavy censoring.

Figure 4. Circular Design Region with $n = 20$ Runs for a Second-order Model in Two Factors.

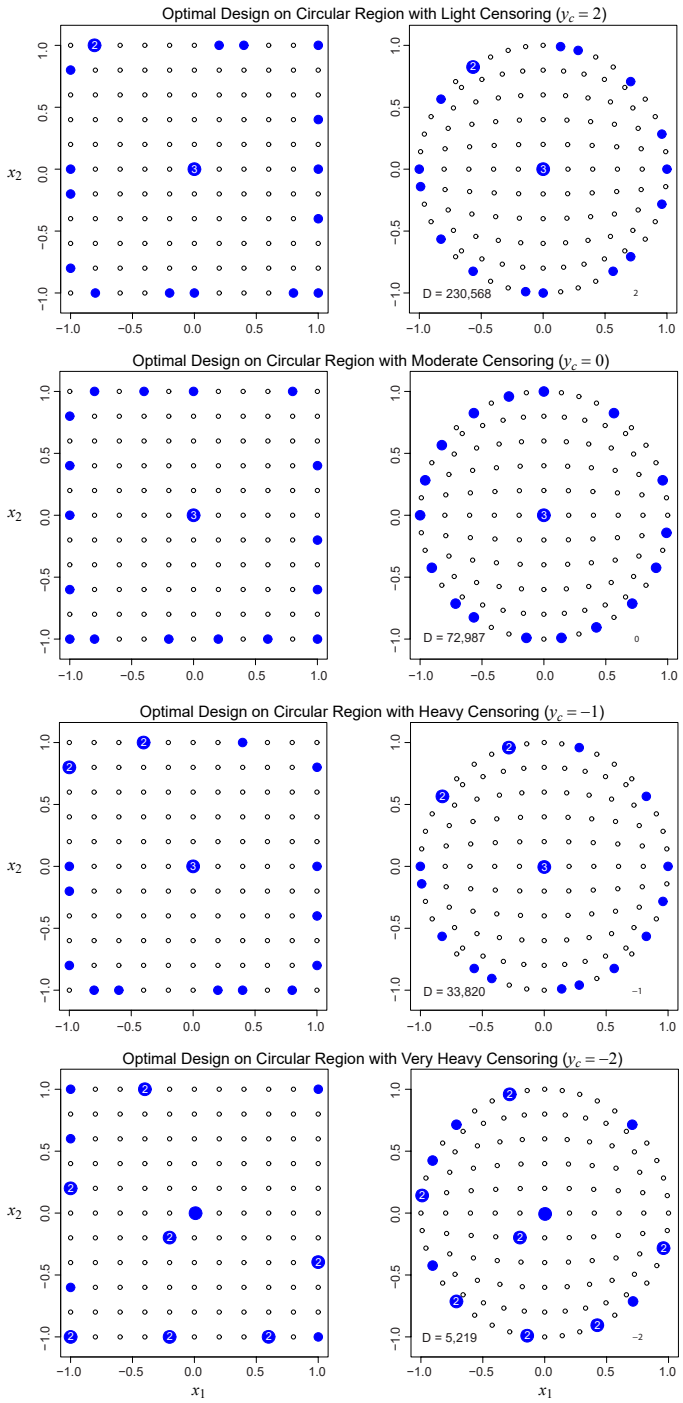


(a) The prior for β_1, β_2, β_{11}, β_{22} and β_{12} is symmetric about the origin.

2.6 Conclusion

Here we developed an approach to perform D-optimal design for life testing models with parameter uncertainty and censoring. We further implemented our process on the 1st order, 1 predictor model through the 2nd order, 2 predictor model. "Fine" tuning is not necessarily needed past the tenth (.1) decimal for providing a high utility of the design. We then expanded the typical grid design region into a trapezoidal or spherical grid. We also varied the amount of censoring when establishing the optimal designs. That is, placing runs of an experiment at the higher end of the grid, i.e. the low stress levels, would increase run times and lead to censoring which provides less information than setting runs at the lower end of the grid.

Figure 5. Optimal Designs for $n = 20$ Runs with Asymmetric Prior That Guarantees an Increasing Link Function in Both x_1 and x_2.



(a) Censoring is light, moderate, heavy and very heavy.

Chapter 3

TEST PLANNING OF REPAIRABLE SYSTEMS VIA POISSON REGRESSION MODELS

3.1 Introduction and Literature Review

A prime question to any system being developed pertains to its reliability, availability, and maintainability. The expected lifetime and mean time between failure (MTBF) helps address these areas of concern. It is well known that reliability estimates using data from developmental testing (DT) are often optimistic estimates of reliability under operationally relevant conditions. This is especially true when dealing with complex systems. Although there is no precise technical definition of a "complex system," most researchers would probably agree that it is a system composed of many interacting parts in a manner such that the collective behavior of those parts together is more than the sum of their individual behaviors (Newman 2011).

One problem in using an optimistic estimates derived during DT is the systems will be viewed ready for operational use when, in fact, their operationally relevant reliability may be substantially smaller. As a result, either testers will find a large number of design defects during operational testing, an outcome inconsistent with the structure of operational testing, or a system with potentially deficient reliability will be promoted to full-rate production, leading to more severe consequences in the future (Council 2014). Design changes at the operational testing phase is rarely an option due to the inefficiency and cost associated with such decision. In other words,

fixing a problem at the later stage of system design is much more expensive than at an earlier stage.

To mature the reliability of a complex system under development, we must detail a reliability growth plan that depicts how the systems reliability is expected to increase over the testing period (Ellner and Hall 2006). Current methods involve selecting a model based on the type of data and the method of analysis. Poisson regression is one such method. Poisson regression models have been applied to range of processes and systems from the medical community to mechanical engineering. Louzada-Neto and Tomazella (2011) used Poisson in exploring tumor growths, Cameron and Trivedi (2013) provides examples of hospital visits and asthma counts and Saldanha, De Simone, and Melo (2001) uses Poisson in modeling water pump failures. This array of application shows the flexibility in Poisson modeling.

Poisson models can represent either homogeneous or non-homogeneous failure/repair process. The Homogeneous Poisson Process (HPP) is a special case of a renewal process which is often used in practice (Aminzadeh 2013). It is one of the simplest point process models when modeling the rate of occurrence of failure/repair events for a repairable system and counts the number of failures during a specific time period. This model comes about when the interarrival times between failures are iid exponentially distributed with an intensity parameter λ. Despite the simplicity of this model, it is widely used for repairable equipment and systems throughout industry. Justification for this comes, in part, from the shape of the empirical Bathtub Curve. Most systems (from complex tools, vehicles, and equipment) spend most of their "lifetimes" operating in the long flat constant repair rate portion of the Bathtub Curve. The HPP is the only model that applies to that portion of the curve, so it is

the most popular model for system reliability evaluation and reliability test planning (Guthrie 2020).

In this chapter we assume failure data of a repairable system follows a Poisson-exponential distribution, i.e., the HPP case, while its intensity parameter has a log-linear relationship with stress factors. We also look at the influence of testing time on designing optimal test plans.

3.1.1 D-Optimal Designs

A design that minimizes the general variance of the regression model coefficients is called the D-optimal design (Montgomery 2019), where D refers that it maximizes the determinant of the Fisher information matrix of all design points within a certain design space. This D-optimality criterion has been used to find designs for Generalized Linear Models (GLMs) by many authors, such as A. C. Atkinson and D. C. Woods (2015), Woods et al. (2006), Yang, Zhang, and Huang (2011), Zhang (2006), and Dobson and Barnett (2018). This design can be generated by an iterative search algorithm such as in the 'OptimalDesign' package within R (Harman, Filova, and Filova 2022).

Optimal designs for GLMs typically require fully specified parametric models, i.e., knowing the specific values of regression coefficients. These designs for regression models have been discussed in Chernoff (1953), Finney et al. (1978) and Abdelbasit and Plackett (1983). However, these "best guess" designs may not be sufficient if the true parameter values are far from the guess. That is, these designs are not as robust when predicted and true parameter values are far apart. Another method is to create a rang of values, or a distribution, for the model parameters. An example of

this is can be found in Zhang and Ye (2014), who investigated the Bayesian design for a one-variable model. The Bayesian D-optimality criterion uses the expected information gain of the Shannon information matrix (Shannon 1948), choosing the design X that maximizes

$$\phi(X) = E[\log(\det I(\beta, X))] \tag{3.1}$$

3.1.2 Poisson Regression model

Lall et al. (2018), Russell et al. (2009), and Wang et al. (2006) have all considered the problem of finding a D-optimal design under a Poisson regression model. A log-linear Poisson regression model is expressed as

$$N_i \sim Poisson(\lambda_i), \quad \text{where } \lambda_i = exp(\mathbf{x_i'}\beta) \tag{3.2}$$

Note that λ_i is the mean of Poisson distribution. In the case of repairable systems, N_i is the number of failures at design point i over a time period of one time unit. We further assume N_i follows a Poisson distribution with mean λ_i, which is equivalent to the HPP with intensity λ_i over a given unit time. We can further define the intensity as a function of predictor factors, x_i, working up to 2 factors and 2nd-order models. Let $\theta = \ln \lambda$, we have

$$\theta_i = \beta_0 + \beta_1 x_i \tag{3.3}$$

$$\theta_i = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 \tag{3.4}$$

$$\theta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \tag{3.5}$$

$$\theta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i} \tag{3.6}$$

$$\theta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} \tag{3.7}$$

For models (3.6) to (3.7), θ_i measures the combined effect of both factors, such as temperature and pressure, with or without interaction or 2nd-order effects. We then find a design space that minimizes the volume of the expected confidence region for a local optimal design, or in terms of Bayesian design, a posterior credible region averaged over the prior distribution.

The general Fisher information matrix for the Poisson regression model can be written as

$$I(\mathbf{X}, \beta) = \sum_{i=1}^{k} p_i \lambda_i (\mathbf{x_i x_i^T}) \tag{3.8}$$

where x_i is a $p \times 1$ design vector of the i^{th} design point, while p is the number of model parameters (or regression coefficients). Note that we use p_i to denote the proportion of whole sample size being assigned to the i^{th} design point, which is different from p, and $\sum_{i=1}^{k} p_i = 1$.

Let the total sample size be n. The Fisher information matrix for the one-factor, first-order Poisson model, Eq. (3.3), can be written as

$$I(\beta_0, \beta_1, X) = n \begin{bmatrix} \sum p_i \lambda_i & \sum p_i \lambda_i x_i \\ \sum p_i \lambda_i x_i & \sum p_i \lambda_i x_i^2 \end{bmatrix} \tag{3.9}$$

For the one-factor, second-order model, Eq. (3.4), the Fisher Information matrix is

$$I(\beta_0, \beta_1, \beta_{11}, X) = n \begin{bmatrix} \sum p_i \lambda_i & \sum p_i \lambda_i x_i & \sum p_i \lambda_i x_i^2 \\ \sum p_i \lambda_i x_i & \sum p_i \lambda_i x_i^2 & \sum p_i \lambda_i x_i^3 \\ \sum p_i \lambda_i x_i^2 & \sum p_i \lambda_i x_i^3 & \sum p_i \lambda_i x_i^4 \end{bmatrix} \tag{3.10}$$

For the two-factor, first-order Poisson model, Eq. (3.5), the Fisher information matrix is

$$I(\beta_0, \beta_1, \beta_2, X) = n \begin{bmatrix} \sum p_i\lambda_i & \sum p_i\lambda_i x_{1i} & \sum p_i\lambda_i x_{2i} \\ \sum p_i\lambda_i x_{1i} & \sum p_i\lambda_i x_{1i}^2 & \sum p_i\lambda_i x_{1i}x_{2i} \\ \sum p_i\lambda_i x_{2i} & \sum p_i\lambda_i x_{1i}x_{2i} & \sum p_i\lambda_i x_{2i}^2 \end{bmatrix} \tag{3.11}$$

For the two-factor interaction model, Eq. (3.6), the Fisher Information matrix is

$$I(\beta_0, \beta_1, \beta_2, \beta_{12}, X) = n \begin{bmatrix} \sum p_i\lambda_i & \sum p_i\lambda_i x_{1i} & \sum p_i\lambda_i x_{2i} & \sum p_i\lambda_i x_{1i}x_{2i} \\ \sum p_i\lambda_i x_{1i} & \sum p_i\lambda_i x_{1i}^2 & \sum p_i\lambda_i x_{1i}x_{2i} & \sum p_i\lambda_i x_{1i}^2x_{2i} \\ \sum p_i\lambda_i x_{2i} & \sum p_i\lambda_i x_{1i}x_{2i} & \sum p_i\lambda_i x_{2i}^2 & \sum p_i\lambda_i x_{1i}x_{2i}^2 \\ \sum p_i\lambda_i x_{1i}x_{2i} & \sum p_i\lambda_i x_{1i}^2 & \sum p_i\lambda_i x_{1i}x_{2i}^2 & \sum p_i\lambda_i x_{1i}^2x_{2i}^2 \end{bmatrix} \tag{3.12}$$

A D-optimal design maximizes the determinate of the Fisher information matrix and, in this case, is maximized when the portion sizes are equal (Wang et al. 2006).

3.2 Applied Poisson Regression Model

For a Poisson regression model with design matrix X and model parameter vector β, the Fisher information matrix is

$$I(X, \beta) = X'WX$$

where $W = \text{diag}(\exp(x_j'\beta))$ is the Hessian matrix. As previously mentioned, we see that the Hessian depends on this unknown parameters. In Wang et al. (2006), they focus on locally D-optimal designs and assume that substantial knowledge about the toxicants under investigation is available. The $X's$ are doses or concentrations of toxicants and are on a continuous scale. Their Poisson regression model is written as

$$y_{ij} \sim Poisson(\lambda_i), \quad \text{where } \lambda_i = exp(x_i'\beta), \tag{3.13}$$

3.2.1 Continuous vs Discrete Design Space

Wang et al. (2006) discusses the D-optimal design for Poisson regression models with a continuous design space. They investigated how the D-optimal designs depend on the model parameters for the one-variable second order model, and two-variable interaction model. In their content of toxicity, the y_{ij} of eq 3.13 is the number of organisms or cells that survive the experiment for the j^{th} replicate at the i^{th} design point. We will explore the one predictor, second-order model.

$$\lambda_i = exp(\beta_0 + \beta_1 x_i + \beta_{11} x_i^2), \tag{3.14}$$

We investigate how their optimal design on a continuous region, compares to one over a discrete design region. We will use the Fischer information matrix in determining the optimal design. For a 1 factor 2nd order model, assuming equal proportions and runs, this matrix becomes:

$$I(X, \beta) = \begin{bmatrix} \sum_{i=1}^{3} \lambda_i & \sum_{i=1}^{3} \lambda_i x_i & \sum_{i=1}^{3} \lambda_i x_i^2 \\ \sum_{i=1}^{3} \lambda_i x_i & \sum_{i=1}^{3} \lambda_i x_i^2 & \sum_{i=1}^{3} \lambda_i x_i^3 \\ \sum_{i=1}^{3} \lambda_i x_i^2 & \sum_{i=1}^{3} \lambda_i x_i^3 & \sum_{i=1}^{3} \lambda_i x_i^4 \end{bmatrix}$$

3.2.1.1 Example

Recall that Wang et al. (2006) used a continuous design space, but here we compare it to a discrete design region with a 0.1 spacing between design points. We follow the same restrictions to design space when comparing the optimal design using the r and c values set by Wang et al. (2006). In short, the r value varies from model to model based on the β parameters from equations 3-7. For example, model 3.4, $r = \frac{\beta_1^2}{\beta_{11}}$ and for model 3.6, $r = \frac{\beta_{12}}{\beta_1 \beta_2}$. The c value is the lower boundary of the survival

proportion. That is, it is a lower restriction to the design space and should be chosen by practitioners and typically will vary from case to case.

Table 5 shows our discrete design points for a single factor, second-order model, Eq. (3.4). Similar to Wang et al. (2006) Table 6, our first point was at the extreme of 1, except for $r = 5$ and $r = 10$. If we look at $r = -5$, we see our design places the lower point at 0.1 and the middle at 0.4. Intuitively, looking at the continuous design space, we simply round each continuous value up to the nearest tenth. However, if we look at $r = -1$, we see that we can't simply round up for the optimal design. The middle point jumps to 0.6 in the discrete case while the continuous case, if rounding, would be at 0.5. As we follow Wang et al. (2006) in both r and c values, we see that differences exists between the optimal designs, which give the indication that the design, continuous or discrete, plays a role in the optimally.

Table 5. Discrete Design Space: One-Toxicant Second-Order Model.

r =	c = 0.01		c = 0.2		c = 0.4	
$\frac{\beta_1^2}{\beta_{11}}$	q_2	q_3	q_2	q_3	q_2	q_3
-50	0.3	0.1	0.6	0.2	0.7	0.4
-20	0.3	0.1	0.6	0.2	0.7	0.4
-10	0.3	0.1	0.7	0.2	0.7	0.4
-5	0.4	0.1	0.7	0.2	0.7	0.4
-1	0.6	0.1	0.7	0.2	0.7	0.4
0	0.6	0.1	0.8	0.2	0.8	0.4
5	0.3^a	0.1^a	0.6	0.2	0.7	0.4
10	0.3^a	0.1^a	0.6	0.2	0.7	0.4
20	0.3	0.1	0.6	0.2	0.7	0.4
50	0.3	0.1	0.6	0.2	0.7	0.4

(a) Upper bound $q_1 = 1$ except for r=5 and r=10 being .8 and .9

In section 3.3 we explore further into how the design space affects the optimal

Table 6. Wang et al. (2006) Continuous Design Space: One-Toxicant Second-Order Model

r =	c = 0.01		c = 0.2		c = 0.4	
$\frac{\beta_1^2}{\beta_{11}}$	q_2	q_3	q_2	q_3	q_2	q_3
-50	0.3052	0.0117	0.5275	0.2000	0.6693	0.4000
-20	0.3304	0.0150	0.5359	0.2000	0.6737	0.4000
-10	0.3594	0.0190	0.5264	0.2000	0.6776	0.4000
-5	0.3964	0.0244	0.5626	0.2000	0.6864	0.4000
-1	0.4959	0.0401	0.6185	0.2000	0.7223	0.4000
0	0.6825	0.0100	0.7540	0.2000	0.8323	0.4000
5	-	-	-	-	0.6210	0.4000
10	-	-	0.4776	0.2000	0.6496	0.4000
20	0.2158	0.0100	0.5048	0.2000	0.6587	0.4000
50	0.2675	0.0100	0.5162	0.2000	0.6633	0.4000

design and how expanding the design space does not necessarily mean the optimal points will expand proportional to the design region.

3.2.2 Accounting for Length of Time

In section 3.1.2 we ignored time and assumed we had an equal time period of 1. Suppose now we need to account for different time periods. We will continue to assume a constant intensity λ. Therefore, we have a system to be tested for a time period t_i. During this testing period, there will be a failure, followed by a repair, and then more testing. Here we will assume the simpler HPP model. Each time-between-failure has the same exponential distribution, so the number of failures has a Poisson distribution. We can treat each failure as from an identical non-repairable system, and the failure times of these non-repairable systems follow an exponential distribution. Lastly, for ease of computations, we end testing with a failure, as we would in the non-repairable

systems case. In following section 3.1.2, we have

$$N_i \sim Poisson(\Delta\lambda_i), \quad \text{where } \lambda_i = exp(\mathbf{x_i'}\beta) \tag{3.15}$$

where Δ is some total test time, $(0, t)$. The likelihood function and the expected value becomes

$$L(\beta) = \prod_{i=1}^{n} \frac{\lambda_i^{N_i} exp^{-\lambda_i}}{N_i!} \tag{3.16}$$

$$\begin{aligned} E[N_i] &= \int_0^{t_i} \lambda_i dz \\ &= t_i\lambda_i - 0\lambda \\ &= t_i\lambda_i \end{aligned} \tag{3.17}$$

Connecting this to the Fisher information matrix, the partial derivatives remain the same but taking the expected value now accounts for total testing time. For example, using the 2nd partial derivative of β_0 without time we have

$$\begin{aligned} \frac{\partial^2 L}{\partial \beta_0^2} &= \sum_{i=1}^{n} \mathrm{e}^{\beta_0 + \beta_1 x_i} \\ &= \sum_{i=1}^{n} \lambda_i \end{aligned} \tag{3.18}$$

where L is the join log-likelihood of the Poisson random variables. If we now take into account testing time, the expected value becomes

$$\begin{aligned} E[\frac{\partial^2 L}{\partial \beta_0^2}] &= \int_0^{t_i} \sum_{i=1}^{n} \lambda_i d(z) \\ &= \sum_{i=1}^{n} t_i\lambda_i \end{aligned} \tag{3.19}$$

Because of the memory-less property, we see that the only difference is a multiple of the time-period, t. Hence if the original time period was 1 hr, or 60 minutes, then

if we wanted to look at between 60 and 90 minutes, our $t = .5$. From here forward, we will simply state t_i as the time period at design point i.

Now, looking at equation 3.3, and ignoring the proportion size p in the information matrix of 3.9, and letting t be the time period, the new information matrix becomes

$$I(\beta_0, \beta_1, X) = \begin{bmatrix} \sum t_i\lambda_i & \sum t_i\lambda_i x_i \\ \sum t_i\lambda_i x_i & \sum t_i\lambda_i x_i^2 \end{bmatrix}. \tag{3.20}$$

similarly for the 2 factor interaction model, eq 3.6 the Fisher Information matrix, when accounting for time periods. becomes

$$I(\beta_0, \beta_1, \beta_2, \beta_{12}, X) = \begin{bmatrix} \sum t_i\lambda_i & \sum t_i\lambda_i x_{1i} & \sum t_i\lambda_i x_{2i} & \sum t_i\lambda_i x_{1i}x_{2i} \\ \sum t_i\lambda_i x_{1i} & \sum t_i\lambda_i x_{1i}^2 & \sum t_i\lambda_i x_{1i}x_{2i} & \sum t_i\lambda_i x_{1i}^2 x_{2i} \\ \sum t_i\lambda_i x_{2i} & \sum t_i\lambda_i x_{1i}x_{2i} & \sum t_i\lambda_i x_{2i}^2 & \sum t_i\lambda_i x_{1i}x_{2i}^2 \\ \sum t_i\lambda_i x_{1i}x_{2i} & \sum t_i\lambda_i x_{1i}^2 x_{2i} & \sum t_i\lambda_i x_{1i}x_{2i}^2 & \sum t_i\lambda_i x_{1i}^2 x_{2i}^2 \end{bmatrix}. \tag{3.21}$$

For an example, if we look at the determinate for the information matrix (3.20), we see that the determinant is

$$det[I] = (\sum t_i\lambda_i)(\sum t_i\lambda_i x_i^2) - (\sum t_i\lambda_i x_i)^2. \tag{3.22}$$

3.2.2.1 Optimality of Time

Wang et al. (2006) implies that the D-optimal design is an equal-allocation design. So equal portions allocated to each of the design points. But unlike the proportion size, there is no restriction on the time periods. However, infinite time is unrealistic and we will be restricted by some total time interval T. Hence dividing out the test time periods, t_i becomes a priority. Because the D-optimal design is equally allocated,

for a non-repairable system, the division is easy and straight forward. The testing time would be equally portioned amongst the design points. That is, for a three-run design, we'd allocate $\frac{1}{3}$ of the total time to each design point because we'd also allocate an equal number of systems to each design point and test to failure. Things become a bit more complicated when dealing with repairable systems.

In the testing of repairable systems, we assume different intensity functions, $\lambda's$, at each design point and therefore assume different number of failures (which can be treated at different systems under test). We now need to balance the equation, i.e. equally allocating the runs/failures, to ensure we maximize the information gained. One way to do this is to treat $t_i\lambda_i$ as a proportion. Since each λ_i is a constant, although possibly different, we would divide the times in a manner such that all $t_i\lambda_i's$ equalled each other. This effectively makes $t_i\lambda_i$ a constant value and ensures the number of failures, i.e. the number of systems under test at each design point, equal. The design is now equally-allocated and therefore we have maximized the information gained and simplifies the information matrix. As an example, the information matrix (3.12), would reduce to

$$I(\beta_1, \beta_2, \beta_{12}, X) = \begin{bmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} & \sum x_{1i}^2x_{2i} \\ \sum x_{1i}x_{2i} & \sum x_{2i}^2 & \sum x_{1i}x_{2i}^2 \\ \sum x_{1i}^2x_{2i} & \sum x_{1i}x_{2i}^2 & \sum x_{1i}^2x_{2i}^2 \end{bmatrix}. \tag{3.23}$$

3.2.2.2 Example: Varying Failure Rate of Occurrence

Consider a test under the "use" condition of the system. We set our x coefficients to zero, 0, and β_0 becomes the final arrival rate of failures. Further consider if the expected values, μ_i, are equal. That is $\mu_1 = \mu_2 = ... = \mu_k$. Then we get the information matrix matrix such that $X^TWX = \mu(X^TX)$. We know the optimal

design is a balanced design. So if we know, or have some idea of the intensity, λ_i, at each design point, and we vary our testing time at each of those design points so that $\lambda_1 t_1 = \lambda_2 t_2 = ... = \lambda_k t_k$, then we can easily design our test as a balanced design and have a reasonable argument as to why this design will be optimal, since it maximizes the information gained.

For simplicity, suppose we have a single factor second order model as in eq. 3.4 for the log-linear Poisson regression. Recall that the intercept β_0 does not effect the optimal design. The information matrix can be seen as

$$I(\beta_0, \beta_1, \beta_{11}, X) = \begin{bmatrix} \sum t_i \lambda_i x_i^2 & \sum t_i \lambda_i x_i^3 \\ \sum t_i \lambda_i x_i^3 & \sum t_i \lambda_i x_i^4 \end{bmatrix} \tag{3.24}$$

Assuming that our feasible design region ranges from [0,1]. The D-optimal design, in regards to partitioning, would be (0, 0.5, 1). Figure 6, generated from the OptimalDesign package within R, displays the exact design by partitioning the runs equally amongst the 3 design points without factoring in testing time at said points.

Now, lets suppose we have these three systems but we plan to make repairs during testing. If we have 170 hours of total testing time, we may reasonably assume to test for equal periods 56.66 hrs at each design point. However, now assume that the design space has accelerating factors and that we expect the stress and failure rate at design point .06 to be two times that of the use condition and design point 1 to be five times at use condition. This implies we have some prior expectation of how the system will behave. Therefore, our intensity is λ at design point 0.1, then we have (λ, 2λ, 5λ) respectively for each of the design points. We then apply time t_i to each of the design points such that $2\lambda t_{0.6} = \lambda t_{0.1} = 5\lambda t_1 \implies 2t_{0.6} = t_{0.1} = 5t_1$. We now wish to optimize the time and discover that the test times at each point are $t_{0.01} = 100$, $t_{0.6} = 50$, $t_1 = 20$ hours. Figure 7 shows how to partition the runs based

on our assumption of λ, the accelerated factor effects, and our total testing time of t=170 hrs. This is valuable to the company conducting the test, especially when the cost of testing at the accelerated factors costs more than at use condition.

Figure 6. Standard Design

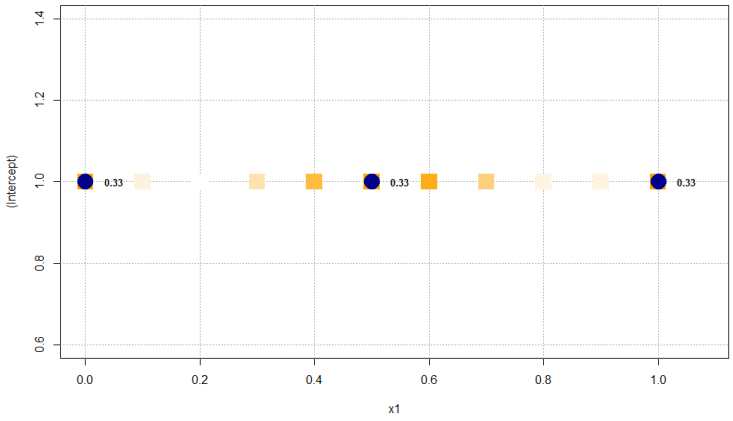
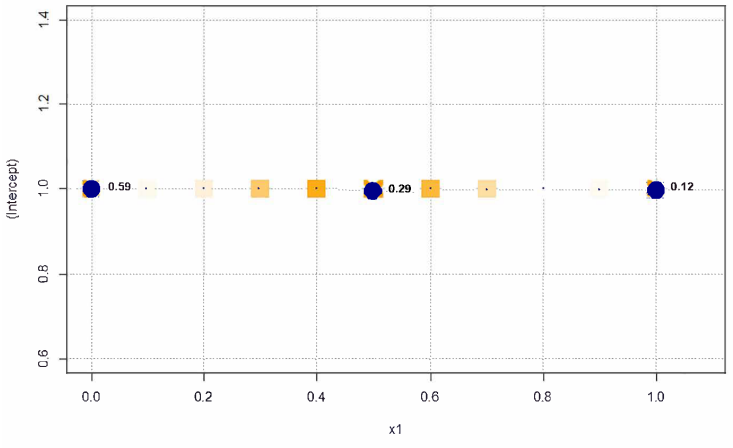


Figure 7. Design Accounting for Time



3.2.2.3 Example: Diminishing Information

Now suppose, based on prior knowledge, that the information we are able to gain at each design point diminishes on a linear scale from 1 to 0.1. That is, for a single factor design, at design point 0, we expect to be able to test all systems to failure, gaining 100% of the expected information, but at design point 1, we expect to gain one tenth of the information needed due to time restrictions. That is, we have a condition on the amount of time we can test along the design space. This may happen for a variety of reasons such as cost constraints associated with testing at the higher stress and the need to limit time and hence expect to end the testing earlier than would be required for system failures.

Figure 8 shows how the design points shift from the locally optimal design at design points (0.0, 0.5, 1.0) to Figure 8 with design points (0.0, 0.6, 1.0). The optimal design is now trying to account for the information lost at design point (1), given reduced allotted times as we approach the higher stress level, thereby shifting the middle design point to the right, closer to (1).

We can further look at if the only restriction was on the highest design point at 1. Suppose at this level, we can only run the test for a tenth (0.1) of the expected need time but all other points we expect the testing to run for as long as needed. Unlike in Figure 8, where the design points shifted to the right, Figure 10 actually shifts the design points to the left. The optimal design is to ignore the highest setting and instead place the design points at 0.0, 0.4, and 0.9. The results are interesting in that depending on how we inject our time variable, the optimal design change may not necessarily be intuitive.

Figure 8. Varying Time and Information Gained: Unrestricted Testing Time

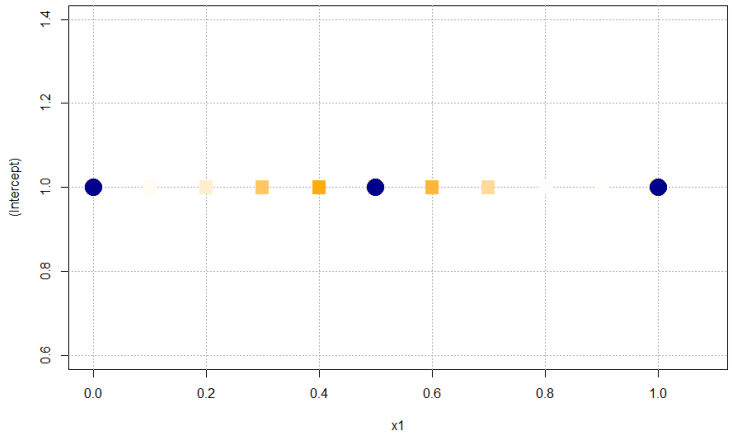
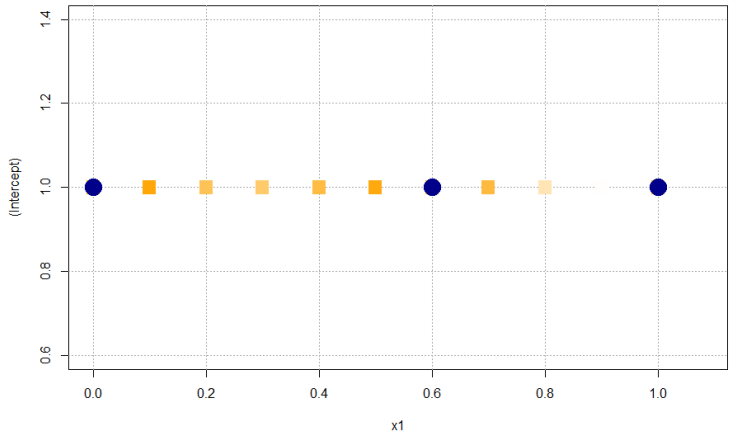


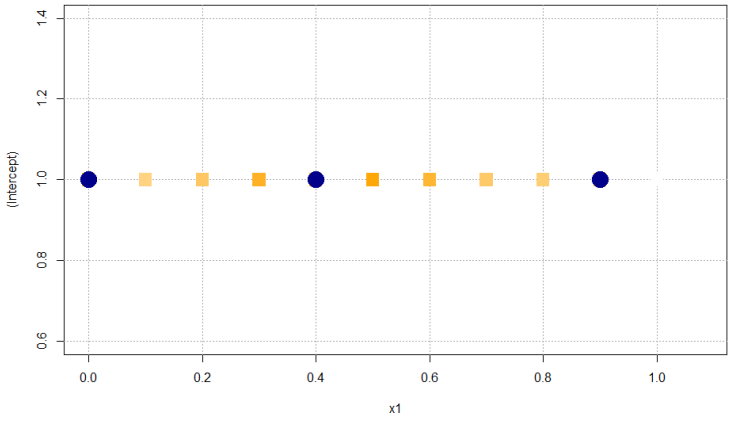
Figure 9. Varying Time and Information Gained: Restricted Testing Time



3.3 Expanding Design Spaces

We may also look at how changing in design space affects the optimal design. Suppose we have two models that follow that of eq 3.3 and also eq.3.6. We will first look at the simple linear 1 predictor model with 2 systems under test and expand the

Figure 10. Varying Time at Design Point 1



design space from [0,1] to [0,4]. The model is as follows

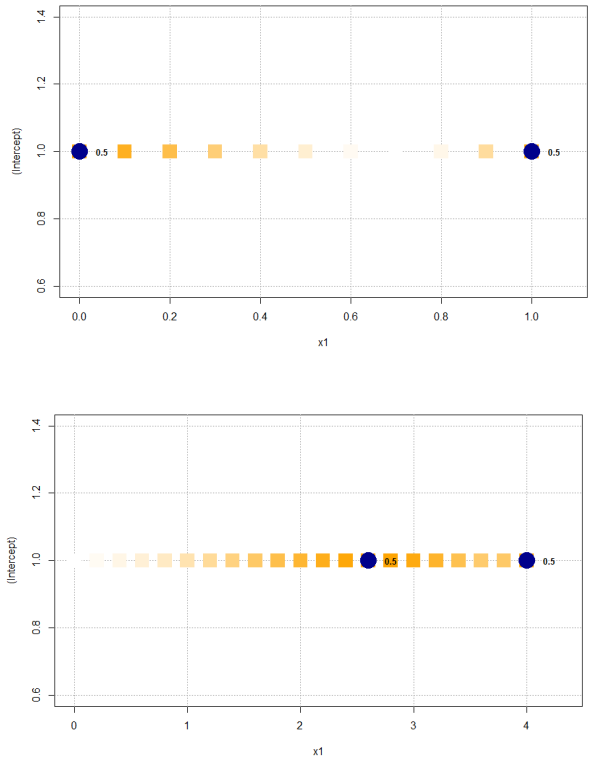
$$ln(\lambda_i) = \theta_i = \beta_0 + \beta_1 x_{1i} \tag{3.25}$$

$$= 1 + 1.5x_{1i} \tag{3.26}$$

3.3.1 Example: 1 Factor Model

Under the assumption of normality, one would assume the runs within the respected design spaces would be at both end points. However, taking a look at the optimal design in Figure 11 we notice this is not the case when assuming the failures arrive as a Poisson counting process. Figure 11, generated from the OptimalDesign package within R, displays the exact design by partitioning the runs equally amongst the 2 design points. Its important to note that the β coefficient is positive in this example, causing the design points to skew towards 4. A negative β coefficient would have caused the design points to be skewed towards 0. The affect is more dramatic as the β coefficient increases as seen in Figure 12

Figure 11. 1 Factor Linear Model, $\beta_1 = 1.5$



3.3.2 Example: 2 Factor Model

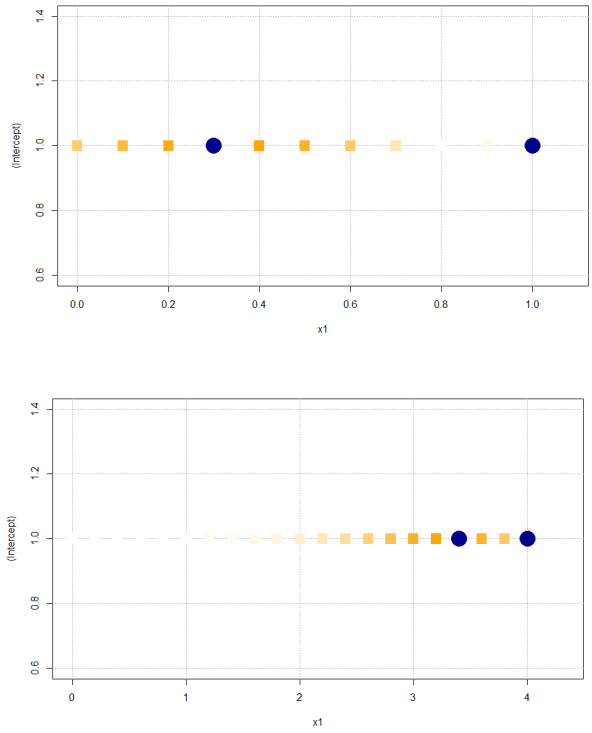
We now look at a 2 factor full interaction model with coefficients as follows:

$$ln(\lambda_i) = \theta_i = 1 + 2x_{1i} + 2x_{2i} - 1x_{1i}^2 - 1.5x_{2i}^2 + 1.5x_{1i}x_{2i} \tag{3.27}$$

Notice the nonlinearity in Figure 13 and similar to the single factor model, by expanding the design space does not necessarily guarantee runs at or even near the end points.

Figures 11 and 13 illustrate the importance of design space regions. The Poisson assigns probability at each design point, based on the factor coefficients, with the optimal design not expanding in a manner seen within the assumptions of normality.

Figure 12. 1 Factor Linear Model, $\beta_1 = 3$



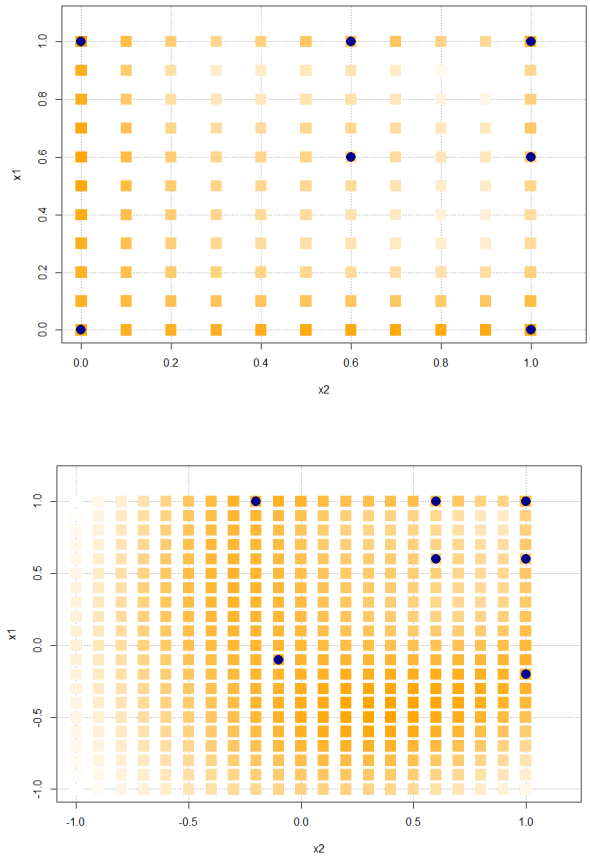
However, a quick look at a graph of the model, helps with visualization and why the design doesn't expand in the normal fashion. Figures 14 illustrates the model behavior in the 3 dimensional space.

A. Atkinson and D. Woods 2015 further discusses designs for generalized linear models for other distributions.

3.4 Bayesian Approach

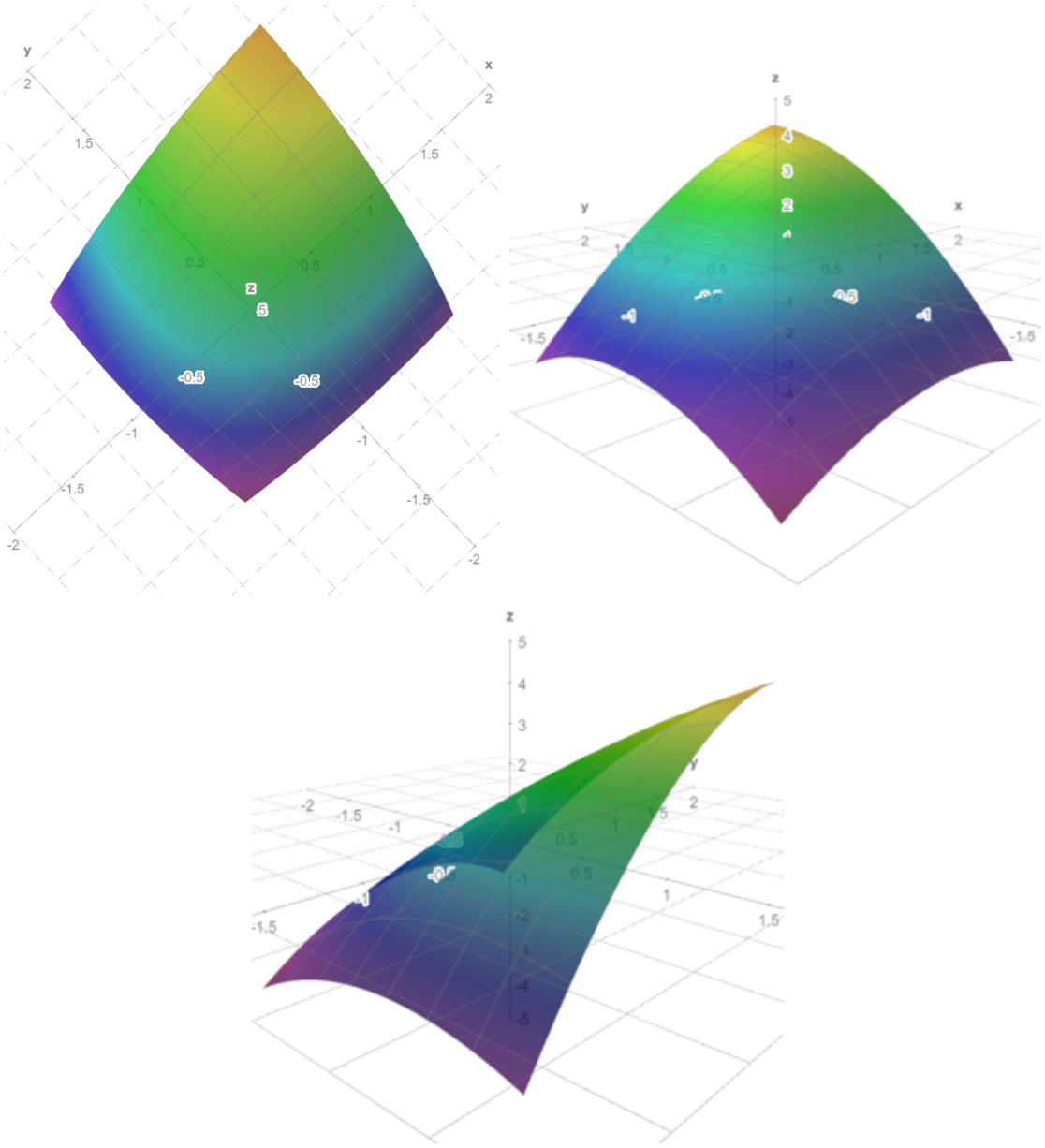
In section 3.2.2 we looked at how time factored into the optimal designs based on known, or expected, coefficients. There are many cases where these β coefficients will be unknown and only a range of values will be expected. A Bayesian approach

Figure 13. Expanding Design Space: 2 Factor Full Order Model



can be used by placing a prior distribution on the parameters thereby leading to an appropriate selection of experimental design. This approach provides a mathematically rigorous and principled based methodology for making decisions under uncertainty, and in contrast to traditional methods, the Bayesian approach itself is very flexible. Taylor et al. (2022) explores using the Bayesian approach on various predictors and ordered models to select appropriate D-optimal designs when the response variable follows a Weibull distribution. A similar, but simpler, method can be applied here when assuming the HPP case.

Figure 14. 3-D Image of the 2 Factor Full Order Model



3.4.1 Example: Estimating Model Parameters

Suppose we have a similar situation as in Figure 9 where we have 3 systems under test and a restricted time space. We now wish to determine the optimal design based

on a range to the β coefficients, instead of fixed predetermined values. For example, suppose

$$N_i|\boldsymbol{x}_i \sim \text{POS}(\lambda = \exp(\beta_0 + \beta_1 x_i + \beta_{11} x_i^2))$$

If we let $Y_i = \log N_i$, then we are effectively assuming

$$Y_i|\boldsymbol{x_i} \sim \text{POS}(\theta = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2). \tag{3.28}$$

Note that if the predictors x_1 and x_{11} are environmental stress variables, which accelerate a test unit's failure process, then such reliability tests are called accelerated life tests (ALTs). In ALT, the value ranges of parameters are typically restrictive, because higher stress levels are supposed to shorten a test unit's lifetime.

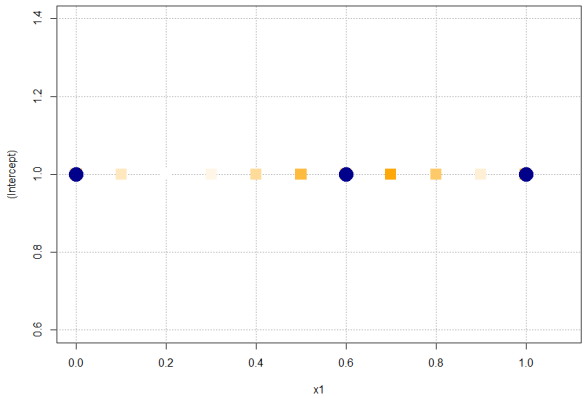
Figure 15 shows two different optimal designs based on possible β range values. For simplicity we selected the design with the greatest utility, however other methods, such as averaging the design over the parameter range may be helpful. Figure 15 provides the optimal design when the assumed coefficient values can range from 0 to 2 by a factor of 0.2; this equates to 27 different combinations the algorithm must run through at each design point. The range was set as

$$\begin{aligned} \beta_0 &\in \text{seq}(0, 1, 0.2) \\ \beta_1 &\in \text{seq}(0, 1, 0.2) \\ \beta_{11} &\in \text{seq}(0, 1, 0.2) \end{aligned}$$

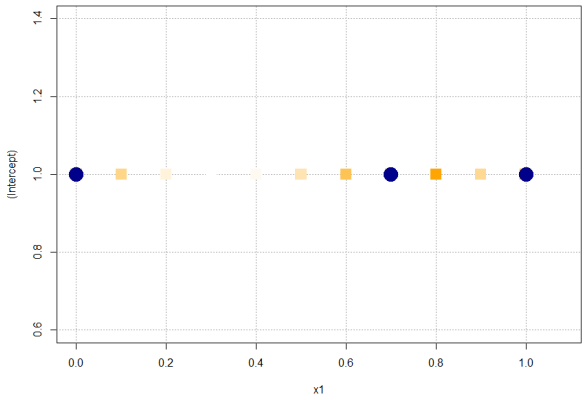
Figure 15a depicts the case when the range is 0 to 2 and equates to 125 different combinations. You can see as the range of potential β values change, so too does the design.

$$\begin{aligned} \beta_0 &\in \text{seq}(0, 2, 0.2) \\ \beta_1 &\in \text{seq}(0, 2, 0.2) \\ \beta_{11} &\in \text{seq}(0, 2, 0.2) \end{aligned}$$

Figure 15. Bayesian Approach - 1 Factor 2nd Order Model



(a) β's range [0,1]



(b) β's range [0,2]

For simplicity, we assumed equal probability as to the values of the coefficients. That is, each model has the same likelihood as the others. We see how the design changes as the coefficient range widens. That is, we see how the design changes when more uncertainty exists on the behaviour of our model and the intensity function. One may further inject complexity into this model by applying a probability matrix to the values, as in the work of Taylor et al. (2022). This would equate to a more precise

prior distribution for the parameters that could be defined from historical lifetime data or testing data from the same or similar products.

3.5 Conclusion

Through simple applications, we explored D-optimality in combining the design space, testing time, and the use of prior knowledge (i.e Bayesian approach). As we stepped through the discreet HPP models, introducing the time factor at varying stress levels, we saw how testing time affects the optimal design. We then moved on-wards in expanding the design space and noticing the nonlinear resulting designs. We further explored the optimal versus classical design approaches by expanding beyond binomial clinical designs and uncovered the possibilities of obtaining greater information that may be important, although not currently needed.

Chapter 4

TEST PLANNING OF REPAIRABLE SYSTEMS VIA THE POWER MODEL

4.1 Introduction and Literature Review

Although HPP is probably the best known and most frequently used model in the field of reliability (Härtler 1989), this model fails when dealing with complex repairable systems. A model of these systems should be capable of expressing reliability growth or deterioration. This implies that the assumption of identically distributed times between failures, as in the HPP case, is not justifiable because the rate of occurrence between failures depends on time. This is where the non-homogeneous Poisson process (NHPP) shines.

4.2 Non-Homogeneous Poisson Process

Engelhardt and Bain (1986) state the NHPP is a mathematical point process characterized by the following properties:

1. The number of failures $N(t)$ in the time interval $[0, t]$ is a non-decreasing and non-negative random integer.
2. The number of failures in disjoint time intervals are independent random integers.
3. There do not occur simultaneous failures.
4. The expectation of time until the first failure is finite.
5. The mean function $m(t) = E[N(t)]$ is strictly increasing with t.
6. The failure intensity $\lambda(t) = m'(t)$ is differentiable.

The NHPP model is completely characterised by the mean function, $m(t)$, or the failure intensity function, $\lambda(t)$ (Härtler 1989).

The Crow-AMSAA model, as discussed in CH1 is a very common model. However, for simplicity, and to connect the work in CH3, our intensity function for the NHPP case can be thought of as

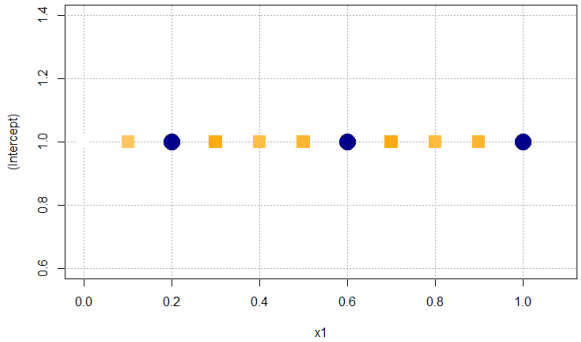
$$\rho_i(t) = t^{\alpha}\lambda_i = t^{\alpha} exp(\mathbf{x_i'}\beta). \tag{4.1}$$

The parameter α is called the Reliability Growth Slope and typical values for growth during reliability improvement tests are in the 0.3 to 0.6 range but grows with any value such that $0 < \alpha < 1$. One may derive a estimator for α as in Guo and Pan (2008), however for our case, we fix α based on previous knowledge. For α values greater than 1, we have deterioration of the system. This usually happens near the end of a systems lifetime but we will assume the beginning life of a new system design and anticipate the system to improve.

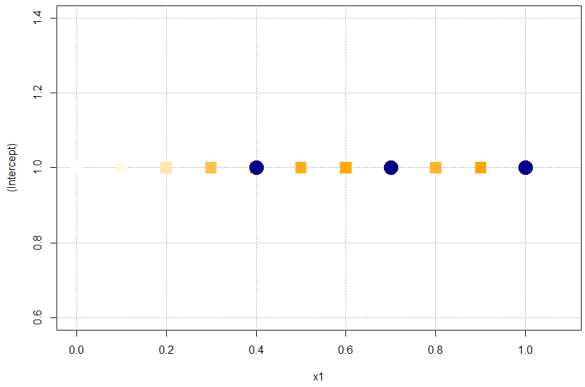
4.2.1 Example: 1 Factor Model

Continuing with a 1 factor 2nd order model, suppose our model follows that of Figure 2(a) but instead of the HPP case, we now allow the intensity to vary with time. Assume we expect to have 1000 hrs of test time, i.e t=1000, and we expect a growth such that $\alpha = 0.3$. Figure 16 shows how this changes the optimal design, shifting the optimal design points to the right. Figure 16a shows how the design changes for $\alpha = 0.6$

Figure 16. NHPP with Design Region [0,1]



(a) NHPP with $\alpha = 0.3$



(b) NHPP with $\alpha = 0.6$

4.2.2 Example: 2 Factor Model

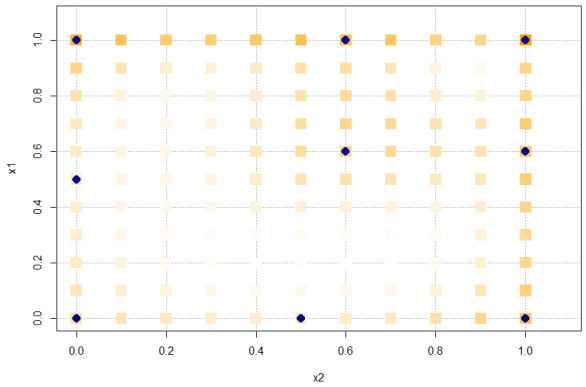
Now suppose we a full 2 factor model and that we have 9 systems we can use for testing. If we assume the model parameters are as follows

$$\lambda_i = exp(1 + 2x_1 + 2x_2 - x_1^2 - 1.5x_2^2 + 1.5x_1x_2), \tag{4.2}$$

then the D-optimal design can be seen in figure 17

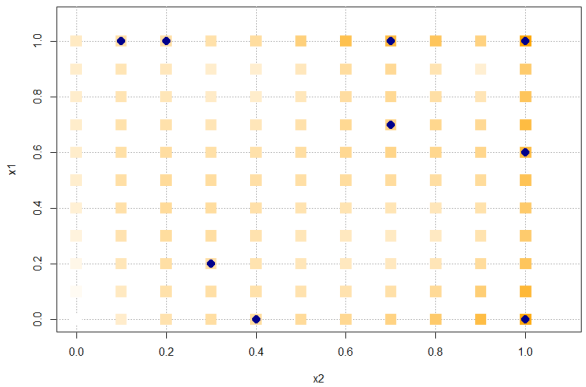
Now assume, similar to a previous example, that we have a linear degradation

Figure 17. Standard D-Optimal Design



in expected information gained (or testing time) as we move right towards the (1,1) level. That is, from (0,0) we expect the full=1 amount of time but as we move right, each design point loses a tenth, 0.1, amount of time such that at (1,1) we have only one tenth information available. Figure 18 shows how the optimal design is changed, shifting runs closer to (1,1) in an effort to accommodate for the loss of information.

Figure 18. Accounting for Time



Although simple examples, we see how accounting for time and using the correct growth slope changes the optimal design. This may help close the gap in the develop-

mental testing reliability predictions and the actual, real-life reliability of the system in the operational environment.

Chapter 5

EXPERIMENTAL DESIGN

5.1 Why Optimal Designs

The most heavily used designs in industry are the "classical designs" (full factorial designs, fractional factorial designs, Latin square designs, Box-Behnken designs, etc.). They are so heavily used because they are optimal in their own right and have served superbly well in providing efficient insight into the underlying structure of industrial processes (Guthrie 2020). Cases do arise when the classical designs do not cover a particular practical situation. According to Guthrie (2020), they include:

1. limitations on the maximum number of runs
2. too many varying levels for factors
3. the assumption of an underlying model is overly complicated

Yet another important reason is to gain additional information, outside the scope of the current need. Such as many cases in the medical field.

The term off-label drug use (OLDU) is used extensively in the medical literature, continuing medical education exercises, and the media. Radley, Finkelstein, and Stafford (2006) reported that in a group of commonly used medications, 21 % of prescriptions were for an off-label use. While another study suggests that 40% - 60% of all prescriptions written in the United States are for off-label uses. Obtaining a new U.S. Food and Drug Administration (FDA) approval for a medication can be costly and time-consuming. Furthermore, physicians have been involved in legal claims due to an adverse reaction related to a medication prescribed for an off-label

use. These reasons alone are enough to warrant the need and desire to obtain the greatest information possible, from a single experiment. A current example is the drug Tirzepatide, a diabetes drug. It is currently inching closer to FDA authorization for weight loss but physicians have the ability to prescribe the drug now. Had greater information been derived from the initial diabetes testing, physicians may be more inclined to prescribe the medication before the FDA authorization, thereby generating short term revenue for the pharmaceutical company or, at a minimum, speeding up the FDA authorization process.

5.1.1 Example: Tumor Medication

Louzada-Neto and Tomazella (2011) present a natural extension of the conventional accelerated failure time model that accommodates recurrent events. They used an animal tumor example that consisted of 23 rats in Group 1 (Treatment) and 25 in Group 2 (Control). The data refers to the number of days on which new tumors occurred for each animal; a given animal may experience any number of tumors. The main objective of analysis was to assess the difference between Groups 1 and 2 with regard to the development of the tumors. Louzada-Neto and Tomazella (2011) determined the estimated hazard rate to be

$$\lambda(t|x) = 1.107 time^{0.017} \exp(-3.179 - 0.743x)]^{1.107}$$

Now, suppose we expected, from past studies of similar drugs, that we needed to test for 6 - 7 months (176 - 205 days) and assume different levels of censoring will take place; that is ending the trial before/after a tumor develops. Furthermore assume the range of model parameters to be that of Table 7; which coincides with the range determined in Louzada-Neto and Tomazella (2011).

Table 7. Weibull Parameters - 95% C.I

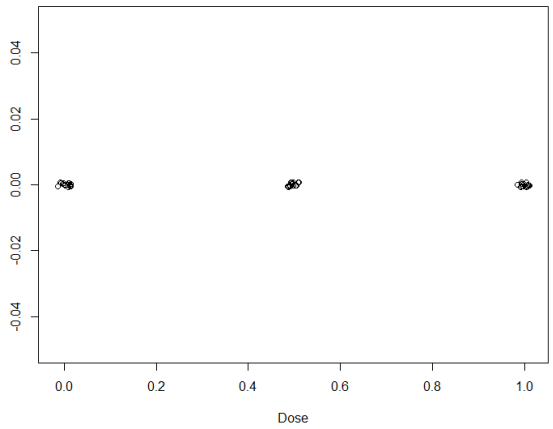
Parameter	MLE	IC (95%)
β	.019	[-0.313; 0.361]
β_0	-4.419	[-4.594; -4.231]

Instead of testing only two groups in the classical DOE manner, one controlled and one at full-dose, we have the ability to test multiple groups at various treatment levels in an optimal way. This possibly gives additional information at results and side-effects from lower doses. If we assume a Weibull failure time model and run the parameters through a coordinate exchange algorithm, as in Taylor et al. (2022), we get the D-optimal designs based on 48 rats (or "runs"). We further account for the level of censoring, that is stopping the test before tumors have appeared on the rats. Figure 19 accounts for little to no censoring and evenly divides the points to 16 runs/rats between (0.0, 0.5, 1) while Figure 19a takes into account heavy censoring and divides the rats into 21 rats at the control group 0, 15 rats at 0.3 and 12 rats at the full-dose 1.0.

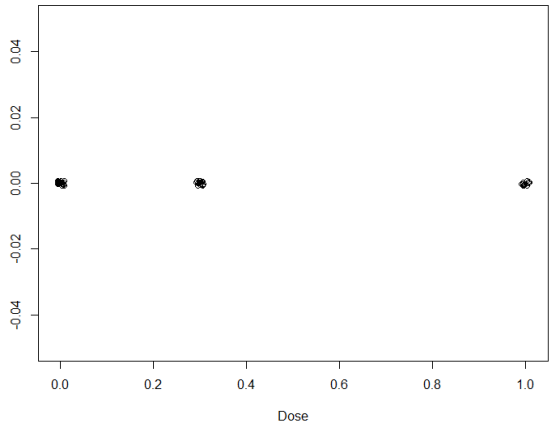
5.2 Methods for Redesigning Experiments

There are two main methodologies when redesigning an experiment; 1) conventional model based design of experiments (MBDoE) and 2) online model-based redesign of experiments (OMBRE). The model based design of experiments (MBDoE) aims at designing a set of experiments that gain the most amount of information in the estimation of some parameters. In our case, designing the experiment in order observe failures and project the reliability of a system. Hypothetically, after each measurement is taken (each failure), one could redesign the remaining part of the

Figure 19. Optimal Design: Rat Dosing Levels



(a) Light Censoring



(b) Heavy Censoring

experiment. However, an excessively small number of experimental data could prove counterproductive (particularly in the early parameter estimations), as it could lead to poor estimators and drive a sub-optimal redesign (Galvanin, Barolo, and Bezzo 2009). The effectiveness of this procedure has been proven and used in a variety of applications such as in Prasad and Vlachos (2008), Galvanin et al. (2009), and

Chakrabarty, Buzzard, and Rundell (2013), but is limited by the fact the design is strongly affected by the experimenters initial guess of the parameters.

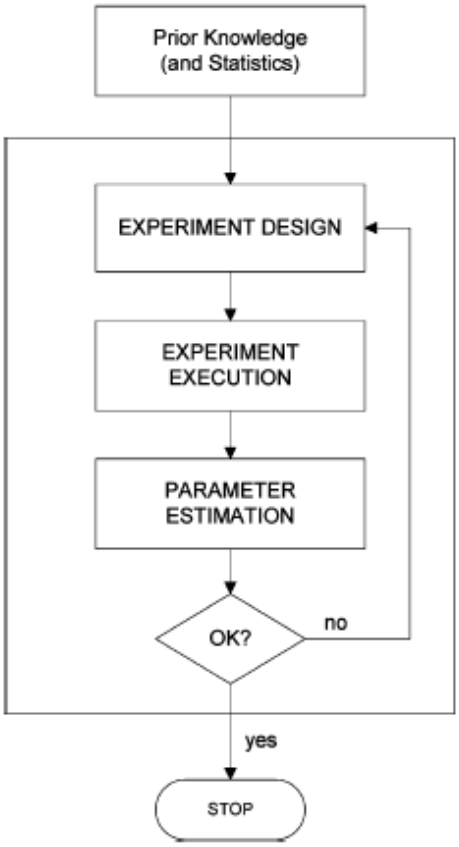
Another method is the online model-based redesign of experiments (OMBRE). In OMBRE, the manipulated variables dynamic profiles and the sampling points allocation are updated by performing one or more intermediate experiment designs (De-Luca, Galvanin, and Bezzo 2016). OMBRE also has limitations that include the redesign policy being decided a-priori by the user without any rational criterion related and it is still affected by the initial parameter uncertainty. Other methods are being explored, such as De-Luca, Galvanin, and Bezzo (2016) information-driven redesign optimization (IDRO), but are not discussed here.

5.2.1 Conventional Model Based Design of Experiments (MBDoE)

The conventional MBDoE (Franceschini and Macchietto 2008) can be described as an iterative loop procedure generally based on the following steps (fig 20):

1. get prior knowledge on the parametric set value (and its related uncertainty);
2. choose an optimal design criterion and design the experiment;
3. start the experiment;
4. at the end of the experimental run, estimate the model parameters;
5. if the desired estimation quality is not reached by the end of the experiment, design a new experiment, based on the model parameters estimated in step 4

Figure 20. MBDoE Procedure



5.2.2 Online Model-Based Redesign of Experiments (OMBRE)

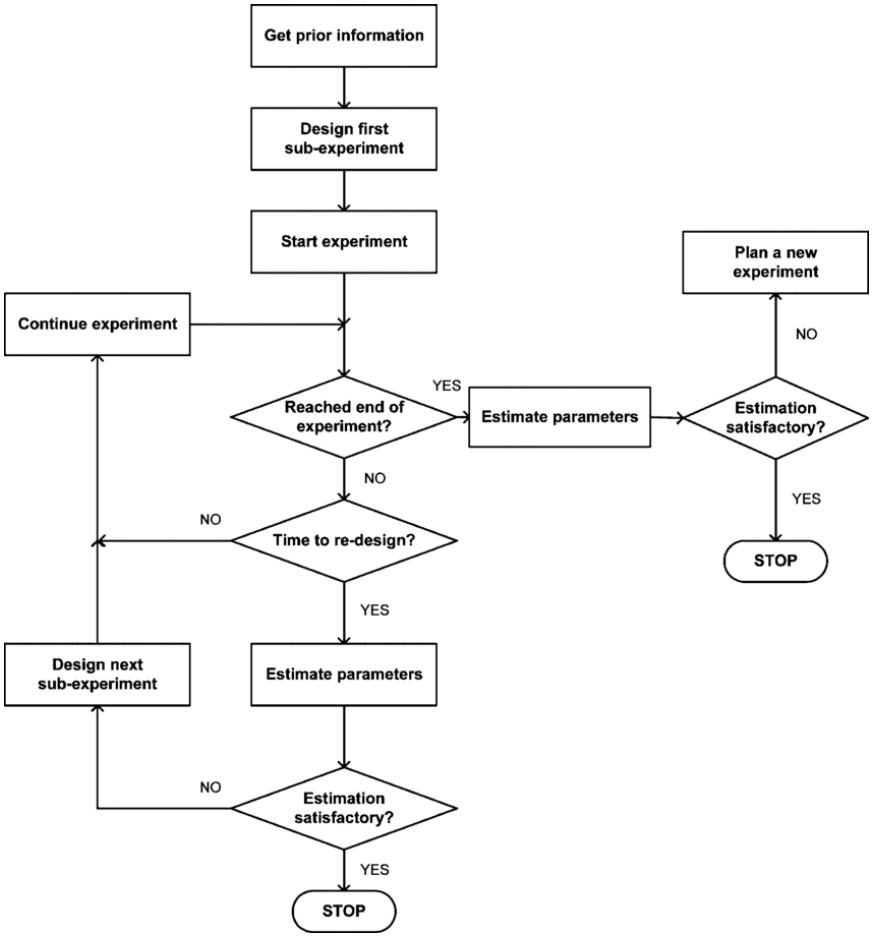
The detailed OMBRE procedure is based on the following steps (fig 21):

1. get prior knowledge on the parametric set value (and its related prior variance-covariance);
2. choose an updating strategy and design the first sub-experiment;
3. start the experiment;
4. if an updating time is reached, estimate the model parameters:
    a. if a statistically sound parameter estimation is achieved, then stop the redesign procedure (and possibly the experiment itself); otherwise

a. redesign the remaining part of the experiment with the update on dynamic information; implement the design in the running experiment, and go to step 4.

5. if a satisfactory parameter estimation is not achieved by the end of the experiment, design a new experiment.

Figure 21. OMBRE Procedure



This chapters main point is two fold. 1) is to justify the need to shift from classical designs into the realm of optimal designs and 2) is to bring to light that even after we have initially designed the experiment, the work is not over. One may want to

redesign throughout, possibly after each testing phase, thereby helping reduce cost and time. All of which is to create a better design bringing forth appropriate failures that will lead to a better, more reliable, system.

Chapter 6

CONCLUSION AND FUTURE WORK

This research began with a focus on applying a Bayesian design to non-repairable systems with a expected systems lifetime following the 2 parameter Weibull distribution. We showed the Bayesian approach to experimental design offers many advantages over other approaches, the most notable of which is the ability to optimize design criteria that are functions of the posterior distribution and that can be easily tailored to the experimenter's design objective (Ryan et al. 2016). We implemented our process on the 1st order, 1 predictor model through the full 2nd order, 2 predictor model with interaction. We noted that "Fine" tuning is not necessarily needed past the tenth (.1) decimal for providing a high utility of the design. We then expanded the typical grid design region into a trapezoidal and spherical grid and accounted for censoring throughout the analysis. All of these features from censoring amounts to design space and parameter estimations changes the D-optimal design.

We then, through simple applications, explored D-optimality in combining the design space, testing time, and the use of prior knowledge (i.e Bayesian approach) to repairable systems with an expected Poisson distribution. We moved from the HPP case to the more complicated NHPP case. In determine the D-optimal design under varying conditions, we showed how the design changes, sometimes not intuitively. These small design changes can have dramatic effects in uncovering failures and thereby effecting predictions to a systems MTBF. We then briefly discussed the optimal versus classical design approaches by expanding beyond binomial clinical

designs and uncovered the possibilities of obtaining greater information that may be important, although not currently needed.

In conclusion, both optimal design of experiment and Poisson models are essential tools for analyzing data in various fields. Combining and utilizing the techniques in this dissertation may lead to more accurate results, improved reliability projections, and may further help researchers gain insights into complex phenomena, to include reliability, maintainability and availability predictions. Future work could expand upon the NHPP case and help further bridge the gap between DT and OT reliability measures and ultimately provide better systems to the consumer and our war-fighters.

REFERENCES

Abdelbasit, Khidir Mohamed, and RL Plackett. 1983. "Experimental design for binary data." *Journal of the American Statistical Association* 78 (381): 90–98.

Activity, U.S. Army Materiel System Analysis. 2011. *Department of Defense Handbook Reliability Growth Management.*

Aminzadeh, MS. 2013. "Bayesian estimation of the expected time of first arrival past a truncated time T: the case of NHPP with power law intensity." *Computational Statistics* 28 (6): 2465–2477.

Asfaw, Zeytu Gashaw, and Bo Henry Lindqvist. 2015. "Unobserved heterogeneity in the power law nonhomogeneous Poisson process." *Reliability Engineering & System Safety* 134:59–65.

Atkinson, Anthony, and David Woods. 2015. "Designs for Generalized Linear Models" (October).

Atkinson, Anthony C, and David C Woods. 2015. "Designs for generalized linear models." *Handbook of design and analysis of experiments* 7:471–514.

Awad, Mahmoud. 2016. "Economic allocation of reliability growth testing using Weibull distributions." *Reliability Engineering & System Safety* 152:273–280.

Bazaraa, Mokhtar S, Hanif D Sherali, and Chitharanjan M Shetty. 2013. *Nonlinear programming: theory and algorithms.* John Wiley & Sons.

Burghaus, Ina, and Holger Dette. 2014. "Optimal Designs for Nonlinear Regression Models with Respect to Non-Informative Priors." *ournal of Statistical Planning and Inference* 154 (1): 12–25.

Cahoon, Joyce, Kate Sanborn, and Alyson Wilson. 2021. "Practical reliability growth modeling." *Quality and Reliability Engineering International* 37 (7): 3108–3124.

Cameron, A Colin, and Pravin K Trivedi. 2013. *Regression analysis of count data.* Vol. 53. Cambridge university press.

Chakrabarty, Ankush, Gregery T Buzzard, and Ann E Rundell. 2013. "Model-based design of experiments for cellular processes." *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 5 (2): 181–203.

Chaloner, Kathryn. 1987. "An approach to experimental design for generalized linear models." *Lecture Notes in Economics and Mathematical Systems* 297:3–12.

Chaloner, Kathryn, and Kinley Larntz. 1989. "Optimal Bayesian design applied to logistic regression experiments." *Journal of Statistical Planning and Inference* 21 (2): 191–208.

———. 1992. "Design of accelerated life testing plans under multiple stresses." *Journal of Statistical Planning and Inference* 33 (2): 245–259.

Chaloner, Kathryn, and Isabella Verdinelli. 1995. "Bayesian Experimental Design: A Review." *Statistical Science* 10 (3): 273–304.

Chernoff, Herman. 1953. "Locally optimal designs for estimating parameters." *The Annals of Mathematical Statistics,* 586–602.

Cobb, George W. 1998. *Introduction to design and analysis of experiments.* Springer.

Council, National Research. 2014. "Reliability Growth: Enhancing Defense System Reliability."

Crow, Larry H. 1975. *Reliability analysis for complex, repairable systems.* Technical report. Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD.

———. 1982. *An improved methodology for reliability growth projections.* Technical report. Army Materiel Systems Analysis Activity, Aberdeen Proving Ground, MD.

DasGupta, Anirban. 1996. "Review of Optimal Bayes Designs." *Handbook of Statistics* 13:1099–1147.

DAVID, S. JOSHUA, and SWETHA. N. 2022. "D OPTIMAL DESIGN FOR POISSON MODEL." *Stochastic Modeling and Applications* 26 (1).

Dette, Holger, and H. M. Neugebauer. 1997. "Bayesian D-Optimal Designs for Exponential Regression Models." *Journal of Statistical Planning and Inference* 60 (2): 331–349.

Dette, Holger et al. 2007. "MAXIMIN AND BAYESIAN OPTIMAL DESIGNS FOR REGRESSION MODELS." *Statistica Sinica* 17 (2): 463–480.

Dobson, Annette J, and Adrian G Barnett. 2018. *An introduction to generalized linear models.* CRC press.

DuMouchel, William, and Bradley Jones. 1994. "A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model." *Technometrics* 36 (1): 37–47.

Ellner, Paul M, and J Brian Hall. 2006. "An approach to reliability growth planning based on failure mode discovery and correction using AMSAA projection methodology." In *RAMS'06. Annual Reliability and Maintainability Symposium, 2006.* 266–272. IEEE.

Engelhardt, Max, and Lee J Bain. 1986. "On the mean time between failures for repairable systems." *IEEE Transactions on Reliability* 35 (4): 419–422.

Escobar, Luis A, and William Q Meeker. 2006. "A Review of Accelerated Test Models." *Statistical Science* 21 (4): 552–577.

Finney, David John, et al. 1978. *Statistical method in biological assay.* Ed. 3. Charles Griffin & Company.

Franceschini, Gaia, and Sandro Macchietto. 2008. "Model-based design of experiments for parameter precision: State of the art." *Chemical Engineering Science* 63 (19): 4846–4872.

Galvanin, Federico, Massimiliano Barolo, and Fabrizio Bezzo. 2009. "Online model-based redesign of experiments for parameter estimation in dynamic systems." *Industrial & Engineering Chemistry Research* 48 (9): 4415–4427.

Galvanin, Federico, Massimiliano Barolo, Sandro Macchietto, and Fabrizio Bezzo. 2009. "Optimal design of clinical tests for the identification of physiological models of type 1 diabetes mellitus." *Industrial & engineering chemistry research* 48 (4): 1989–2002.

Gaver, Donald P, and Patricia A Jacobs. 2014. "Reliability growth by failure mode removal." *Reliability Engineering & System Safety* 130:27–32.

Guo, Huairui, and Rong Pan. 2008. "On determining sample size and testing duration of repairable system test." In *2008 Annual Reliability and Maintainability Symposium,* 120–125. IEEE.

Guthrie, William F. 2020. "Nist/sematech e-handbook of statistical methods (NIST handbook 151)." *National Institute of Standards and Technology: Gaithersburg, MD, USA.*

Harman, Radoslav, Lenka Filova, and Maintainer Lenka Filova. 2022. *Package'OptimalDesign'.*

Härtler, G. 1989. "The nonhomogeneous Poisson process—a model for the reliability of complex repairable systems." *Microelectronics Reliability* 29 (3): 381–386.

Hodkiewicz, Melinda, and Mark Tien-Wei Ho. 2016. "Cleaning historical maintenance work order data for reliability analysis." *Journal of Quality in Maintenance Engineering.*

Hong, Yili, Caleb King, Yao Zhang, and William Q Meeker. 2015. "Bayesian life test planning for log-location-scale family of distributions." *Journal of Quality Technology* 47 (4): 336–350.

"https://www.xarg.org/2017/07/how-to-map-a-square-to-a-circle/." Accessed April 30, 2022.

Kunsong, LIN, and CHEN Yunxia. 2021. "Two new multi-phase reliability growth models from the perspective of time between failures and their applications." *Chinese Journal of Aeronautics* 34 (5): 341–349.

Lall, Shwetank, Seema Jaggi, Eldho Varghese, Cini Varghese, and Arpan Bhowmik. 2018. "D-optimal designs for exponential and Poisson regression models." *Journal of the Indian Society of Agricultural Statistics* 72:27–32.

Lindley, Dennis V. 1956. "On a measure of the information provided by an experiment." *The Annals of Mathematical Statistics* 27 (4): 986–1005.

Louzada-Neto, Francisco, and Vera Lucia Damasceno Tomazella. 2011. "REGRESSION MODELS FOR POISSON DATA BASED ON ACCELERATED LIFETIME TESTING." *Journal of Applied Statistical Science* 19 (2): 183.

De-Luca, Riccardo, Federico Galvanin, and Fabrizio Bezzo. 2016. "A methodology for direct exploitation of available information in the online model-based redesign of experiments." *Computers & Chemical Engineering* 91:195–205.

Meyer, Ruth K, and Christopher J Nachtsheim. 1995. "The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs." *American Statistical Association* 37 (1): 60–69.

Monroe, Eric M, and Rong Pan. 2008. "Experimental design considerations for accelerated life tests with nonlinear constraints and censoring." *Journal of Quality Technology* 40 (4): 355–367.

Monroe, Eric M, Rong Pan, Christine M Anderson-Cook, Douglas C Montgomery, and Connie M Borror. 2010. "Sensitivity analysis of optimal designs for accelerated life testing." *Journal of Quality Technology* 42 (2): 121–135.

Montgomery, D.C. 2019. *Design and Analysis of Experiments, 10e Enhanced eText with Abridged Print Companion.* Wiley. https://books.google.com/books?id=TpK9wAEACAAJ.

Muralidharan, K. 2002. "Reliability inferences of modulated power-law process# i." *IEEE Transactions on Reliability* 51 (1): 23–26.

Nasir, Ehab, and Rong Pan. 2017. "Simulation-based Bayesian optimal design for multi-factor accelerated life tests." *Communications in Statistics-Simulation and Computation* 46 (2): 980–993.

Nassar, Mazen, Sanku Dey, and Saralees Nadarajah. 2021. "Reliability analysis of exponentiated Poisson-exponential constant stress accelerated life test model." *Quality and Reliability Engineering International* 37 (6): 2853–2874.

Newman, Mark EJ. 2011. "Complex systems: A survey." *arXiv preprint arXiv:1112.1440.*

Overstall, Antony, and David Woods. 2015. "The approximate coordinate exchange algorithm for Bayesian optimal design of experiments." *Unpublished.*

Overstall, Antony, David Woods, and Maria Adamou. 2017. "acebayes: An R package for Bayesian optimal design of experiments via approximate coordinate exchange." *arXiv preprint arXiv:1705.08096.*

Pan, Rong, Tao Yang, and Kangwon Seo. 2015. "Planning constant-stress accelerated life tests for acceleration model selection." *IEEE Transactions on Reliability* 64 (4): 1356–1366.

Park, Jung-Won, and Bong-Jin Yum. 1996. "Optimal design of accelerated life tests with two stresses." *Naval Research Logistics* 43 (6): 863–884.

Peng, Yizhen, Yu Wang, YanYang Zi, Kwok-Leung Tsui, and Chuhua Zhang. 2017. "Dynamic reliability assessment and prediction for repairable systems with interval-censored data." *Reliability Engineering & System Safety* 159:301–309.

Prasad, Vinay, and Dionisios G Vlachos. 2008. "Multiscale model and informatics-based optimal design of experiments: application to the catalytic decomposition

of ammonia on ruthenium." *Industrial & engineering chemistry research* 47 (17): 6555–6567.

Radley, David C, Stan N Finkelstein, and Randall S Stafford. 2006. "Off-label prescribing among office-based physicians." *Archives of internal medicine* 166 (9): 1021–1026.

Rasch, Dieter, Jurgen Pilz, Leon R Verdooren, and Albrecht Gebhardt. 2011. *Optimal experimental design with R.* CRC Press.

Rigdon, Steven E. 2002. "Properties of the Duane plot for repairable systems." *Quality and Reliability Engineering International* 18 (1): 1–4.

Rigdon, Steven E, Rong Pan, Douglass C Montgomery, and Laura J Freeman. 2022. *Design of Experiments for Reliability Achievement.* Wiley.

Roy, Soumya, and Biswabrata Pradhan. 2017. "Bayesian optimum life testing plans under progressive Type-I interval censoring scheme." *Quality and Reliability Engineering International* 33 (8): 2727–2737.

Russell, Kenneth G, David C Woods, SM Lewis, and JA Eccleston. 2009. "D-optimal designs for Poisson regression models." *Statistica Sinica,* 721–730.

Ryan, Elizabeth G., Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. 2016. "A Review of Modern Computational Algorithms for Bayesian Optimal Design." *International Statistical Review* 84 (1): 128–154.

Ryan, T.P. 2013. *Sample Size Determination and Power.* Wiley Series in Probability and Statistics. Wiley. https://books.google.com/books?id=WBSwhw-j1ZoC.

Saldanha, Pedro LC, Elaine A De Simone, and PF Frutuoso e Melo. 2001. "An application of non-homogeneous Poisson point processes to the reliability analysis of service water pumps." *Nuclear engineering and design* 210 (1-3): 125–133.

Shah, Kirti R, and BIKAS Sinha. 2012. *Theory of optimal designs.* Vol. 54. Springer Science & Business Media.

Shannon, Claude E. 1948. "A mathematical theory of communication." *The Bell system technical journal* 27 (3): 379–423.

Smucker, Byran J, Willis Jensen, Zichen Wu, and Bo Wang. 2017. "Robustness of classical and optimal designs to missing observations." *Computational Statistics & Data Analysis* 113:251–260.

StatEase. 2020. *Design-Expert > Tutorials > Optimal Design.* Accessed December 23, 2019. https://www.statease.com/docs/v11/tutorials/optimal-numeric/.

Tang, Yincai, Jiayu Fu, Wenchen Liu, and Ancha Xu. 2017. "Bayesian analysis of repairable systems with modulated power law process." *Applied Mathematical Modelling* 44:357–373.

Taylor, Dustin, Steven E Rigdon, Rong Pan, and Douglas C Montgomery. 2022. "Bayesian D-optimal design for life testing with censoring." *Quality and Reliability Engineering International.*

Thomas, Edward V. 2015. *A Statistical Perspective on Highly Accelerated Testing.* Technical report. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).

Tian, Qinglong, Colin Lewis-Beck, Jarad Niemi, and William Meeker. 2022. "Specifying Prior Distributions in Reliability Applications." *arXiv preprint arXiv:2204.06099.*

Wang, Yanping, Raymond H Myers, Eric P Smith, and Keying Ye. 2006. "D-optimal designs for Poisson regression models." *Journal of statistical planning and inference* 136 (8): 2831–2845.

Woods, David C, Susan M Lewis, John A Eccleston, and KG Russell. 2006. "Designs for generalized linear models with several variables and model uncertainty." *Technometrics* 48 (2): 284–292.

Xu, Ancha, and Yincai Tang. 2015. "Reference optimality criterion for planning accelerated life testing." *Journal of Statistical Planning and Inference* 167:14–26.

Yang, Min, Bin Zhang, and Shuguang Huang. 2011. "Optimal designs for generalized linear models with multiple design variables." *Statistica Sinica,* 1415–1430.

Yang, YOU, JIN Guang, PAN Zhengqiang, and GUO Rui. 2021. "An iterated local coordinate-exchange algorithm for constructing experimental designs for multi-dimensional constrained spaces." *Journal of Systems Engineering and Electronics* 32 (5): 1212–1220.

Zhang, Xiangxiang, Jun Yang, and Xuefeng Kong. 2021. "Optimal design of accelerated life tests with two stresses." *Quality and reliability engineering international* 37 (1): 60–77.

Zhang, Ying. 2006. "Bayesian D-optimal design for generalized linear models." PhD diss., Virginia Tech.

Zhang, Ying, and Keying Ye. 2014. "Bayesian D-optimal designs for Poisson regression models." *Communications in Statistics-Theory and Methods* 43 (6): 1234–1247.

Zhu, Yada, and Elsayed A. Elsayed. 2013. "Design of accelerated life testing plans under multiple stresses." *Naval Research Logistics* 60 (6): 468–478.

Zohbi, Ibrahim, Mohieldin Wainakh, and Hamzeh Arafeh. 2016. "Bayesian D-optimal Designs for Weibull Distribution with Censoring." *Quality and Reliability Engineering International* 32 (8): 2789–2802.