

## Designing Degradation Experiments Using a Log-Logistic Distribution

Rana Azmi Dandis and Mohammed Al-Haj Ebrahim

Department of Statistics, Yarmouk University, Irbid, Jordan

**Abstract:** Assessing the reliability of a product is very important to improve the product's quality and to get the trust of customers. Degradation experiments are usually used to assess the reliability of highly reliable products, which are not expected to fail under the traditional life tests. Several decision variables, such as the sample size, the inspection frequency, and the termination time, have a direct influence on the experimental cost and the estimation precision of lifetime information. This paper deals with the optimal design of a degradation experiment where the degradation rate follows a log-logistic distribution. Under the constraint that the total experimental cost does not exceed a predetermined budget, the optimal decision variables are obtained by minimizing the mean squared error of the estimated  $100p^{th}$  percentile of the lifetime distribution of the product. A simulation study and a real example of drug potency data are provided to illustrate the proposed method.

**Zusammenfassung:** Die Beurteilung der Zuverlässigkeit eines Produktes ist sehr wichtig, um dessen Qualität zu verbessern und um das Vertrauen der Kunden zu erhalten. Degradationsexperimente werden in der Regel verwendet, um die Zuverlässigkeit der höchst-zuverlässigen Produkte zu beurteilen, von denen man nicht erwartet, dass diese unter den traditionellen Tests überhaupt fehlschlagen. Mehrere Entscheidungsvariablen wie die Probengröße, die Prüffrequenz, und der Beendigungszeitpunkt, haben einen direkten Einfluss auf die Kosten und auf die experimentelle Schätzgenauigkeit der Lebensdauerinformationen. Diese Arbeit befasst sich mit der optimalen Gestaltung von Degradationsexperimenten, bei denen die Abbaurate einer log-logistischen Verteilung genügt. Unter der Bedingung, dass die gesamten experimentellen Kosten ein vorgegebenes Budget nicht übersteigen, werden die optimalen Entscheidungsvariablen durch die Minimierung des mittleren quadratischen Fehlers des geschätzten  $100p$ -te Perzentil der Lebensdauerverteilung des Produktes erhalten. Eine Simulationsstudie und ein reales datenbeispiel einer Dosis-/Wirkungsstudie werden bereitgestellt, um das vorgeschlagene Verfahren zu illustrieren.

**Keywords:**  $100p^{th}$  Percentile, Degradation Experiments, Highly Reliable Products, Log-Logistic Distribution, Mean Squared Error, Optimal Design, Reliability.

### 1 Introduction

Reliability is an important characteristic of a product's quality. Therefore, assessing the reliability of a product is very important to improve the product's quality and to get the

trust of customers. However, for highly reliable products that are designed to operate without failure for several years, assessing lifetimes using traditional life tests that record only time-to-failure is a difficult task, since no failures are expected to occur over a reasonable amount of time. In such cases, if there exist quality characteristics whose degradation over time is related to failure, then collecting degradation data can be useful in assessing product's lifetime information. Lu and Meeker (1993) discussed statistical methods for using degradation measures to estimate a time-to-failure distribution for a large class of degradation models. In their book about reliability, Meeker and Escobar (1998, Chapter 13) provided various approaches that can be used to assess reliability information based on degradation data.

To conduct a degradation experiment, the experimenter has to pre-specify a critical level of degradation, obtain measurements of degradation at different fixed times, and define that failure occurs when the degradation reaches that level. On the other hand, degradation experiments are expensive, so it is essential to plan them carefully by determining the optimal setting of the three decision variables (i.e. the sample size, the inspection frequency, and the termination time of the experiment) that have a direct influence on the estimation precision and experimental total cost. The total cost of a degradation experiment can be divided into three parts: the cost of tested units, the cost of operating the experiment, and the cost of inspections (Wu and Chang, 2002).

Usually the increase of the sample size and the inspection frequency will increase the precision of the estimates, but this will increase the total cost of the experiment as well. Therefore, an optimal degradation experiment design is the one that can get a tradeoff between estimation precision and budget.

There is an extensive literature on the design of degradation experiments, which shows the importance of the degradation rate distribution of products. Weibull and log-normal distributions were used to describe degradation rates of highly reliable products. For example, Yu and Tseng (1999) studied the optimal setting of the decision variables for a degradation experiment with a linearized degradation model where the degradation rate follows a log-normal distribution. Their study was conducted under the constraint that the total experimental cost does not exceed a predetermined budget and by using the criterion of minimizing the variance of the estimated  $100p^{th}$  percentile of a product's lifetime distribution. Later, Yu and Tseng (2004) investigated the optimal combination of the decision variables for a degradation experiment, but this time with different degradation rate, which is assumed to follow a reciprocal Weibull distribution. They also highlighted the importance of choosing the correct degradation rate for the optimal test plan and showed that an incorrect choice may lead to serious bias.

For very-highly reliability products, the degradation of the quality characteristic is usually very slow and thus it is impossible to have a precise estimation within a reasonable amount of testing time. In such cases, we can conduct an accelerated degradation test (ADT) by using higher stresses to accelerate degradation at each run of the experiment (Meeker and Escobar, 1998). Yu (2002a) addressed the problem of designing an ADT to obtain an efficient interval estimation of mean-time-to-failure and determine the optimal decision variables by minimizing the expected width of its confidence interval, assuming that the degradation rate follows a log-normal distribution. Similarly, but by using the criterion of minimizing the mean-squared error of the estimated  $100p^{th}$  percentile of the

product's lifetime distribution, Yu (2003, 2002b) designed accelerated degradation experiments with log-normal and reciprocal Weibull degradation rates, respectively. For more details see Al-Haj Ebrahim, Alodat, and Arman (2009a); Al-Haj Ebrahim, Eidous, and Kamil (2009b).

Recently the log-logistic distribution has been found to be suitable for describing the degradation rate for highly reliable products (Chiodo and Mazzanti, 2004). In addition, it is often used to analyze lifetime data, especially for events whose rate increases initially and decreases later.

In most practical problems, interest focuses on finding the  $100p^{th}$  percentile of the time to failure distribution  $q_p$ . A commonly used criterion to measure the estimation precision of the  $100p^{th}$  percentile of lifetime distribution is its mean squared error (MSE). Therefore, the optimal values of the decision variables are those that minimize the mean squared error of the estimated  $100p^{th}$  percentile under the constraint the total cost does not exceed a predetermined budget.

The primary objective of this research paper is to determine the optimal setting of the decision variables that affect the experimental cost and the estimation precision of the  $100p^{th}$  percentile of the lifetime distribution of the products whose degradation rates follow the log-logistic distribution. This can be achieved by minimizing the mean squared error of the estimated  $100p^{th}$  percentile under the constraint that the total experimental cost does not exceed a predetermined budget.

The rest of this paper is organized as follows. Section 2 addresses the optimization problem and describes the degradation model and its assumptions. Section 3 explains the framework for solving the optimization problem. Section 4, discusses the results of a simulation study to illustrate the proposed method. A real example of drug potency data is provided in section 5. Conclusions of this study are presented in Section 6.

## 2 The Optimization Problem

In order to understand the problem, we have first to define the decision variables. Thus, let  $n$  be the number of tested units. For each unit let  $f$  denotes the interval between measurements (inspection frequency) and let  $l$  be the number of inspections. Then the termination time is  $(l - 1)f$ . In designing the optimal degradation experiment, the experimenter faces the following decision problems:

- How many devices ( $n$ ) should be tested?
- How to determine the appropriate inspection frequency ( $f$ )?
- How many measurements ( $l$ ) should be taken to collect the data?
- What is the appropriate termination time for the experiment?

Thus, the optimal degradation design problem consists of finding  $(f, l, n)$  that minimizes the mean squared error of the estimated  $100p^{th}$  percentile. However, the determination of  $(f, l, n)$  is restricted to the budget of experiment, say,  $C_b$ . Hence, the optimal design problem can be formulated as follows:

$$\begin{aligned} \text{minimize } \text{MSE}(\hat{q}_p(f, l, n)) \quad \text{subject to } TC(f, l, n) \leq C_b, \\ f, l, n \in \mathbb{N}, l \geq 2, f \geq 1, \end{aligned} \quad (1)$$

where  $\hat{q}_p(f, l, n)$  is an estimator of  $q_p$ ,  $TC(f, l, n)$  is the total cost of conducting a degradation experiment, and  $C_b$  denotes the budget of the experiment.

## 2.1 Degradation Model and Assumptions

Consider the following degradation model

$$y_{ij} = g(t_j, \beta_i) + \epsilon_{ij} = \beta_i t_j + \epsilon_{ij}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, l, \quad (2)$$

where  $y_{ij}$  is the observed degradation measurement of the  $i^{\text{th}}$  unit at time  $t_j$ ,  $g(t_j, \beta_i)$  is the actual degradation path,  $\beta_i > 0$  is a random effect that varies from unit to unit,  $\epsilon_{ij}$  is the random measurement error term,  $n$  is the number of tested units and  $l$  is the total number of inspections for each unit.

Assume that a degradation experiment is conducted under the following conditions:

1. There are  $n$  experimental units randomly selected for conducting a degradation experiment in a specific homogenous environment (e.g. the same temperature, pressure, humidity, etc).
2. The measurements are taken every  $f$  units of time (e.g.  $f$  hours or  $f$  days), until time  $t_l = (l - 1)f$ , where  $l$  is the number of inspections.
3. The error term  $\epsilon_{ij}$  follows a normal distribution with mean zero and variance  $\sigma_\epsilon^2$ .
4. The random effect  $\beta_i$  is assumed to follow a log-logistic distribution with scale parameter  $a$  and shape parameter  $b$  (which is denoted by  $\beta_i \sim \text{Log-logistic}(a, b)$ ).

Note that  $\log \beta_i$  follows the logistic distribution with location parameter  $m$  and scale parameter  $u$ , where  $m = \log a$  and  $u = 1/b$  (which is denoted by  $\log \beta_i \sim \text{Logistic}(m, u)$ ).

The products lifetime  $T$  is defined as the time when the actual degradation path  $g$  crosses the critical level  $D$ . Thus, from the model in (2),  $T$  can be expressed as

$$T = \frac{D}{\beta_i}. \quad (3)$$

Since  $\beta_i \sim \text{Log-logistic}(a, b)$ , it can be easily shown that  $T$  also follows a log-logistic distribution with scale parameter  $D/a$  and shape parameter  $b$  (i.e.  $T \sim \text{Log-logistic}(D/a, b)$ ).

The  $100p^{\text{th}}$  percentile  $q_p$  of the above log-logistic distribution can be expressed as

$$q_p = \frac{D}{a} \left( \frac{p}{1-p} \right)^{1/b}. \quad (4)$$

In addition, we can rewrite the  $100p^{\text{th}}$  percentile in terms of the logistic distribution parameters  $m$  and  $u$  by taking the logarithm on both sides, as follows

$$q_p = \exp \left( \log D - m + u \log \frac{p}{1-p} \right). \quad (5)$$

## 3 The Optimal Design

The following subsections illustrate a framework for solving the optimization problem

### 3.1 Estimation of $q_p$

For  $j = 1, 2, \dots, l$  and based on the observations  $y_{ij}$ ,  $j = 1, \dots, l$ , the least squares estimator (LSE)  $\hat{\beta}_i$  of  $\beta_i$ , conditional on  $\beta_i$ , can be obtained by minimizing

$$\sum_{j=1}^l \epsilon_{ij}^2 = \sum_{j=1}^l (y_{ij} - \beta_i t_j)^2.$$

Clearly, the LSE of  $\beta_i$  is given by

$$\hat{\beta}_i = \frac{\sum_{j=1}^l y_{ij} t_j}{\sum_{j=1}^l t_j^2}. \quad (6)$$

Thus, we have

$$\hat{\beta}_i | \beta_i \sim N \left( \beta_i, \frac{\sigma_\epsilon^2}{\sum_{j=1}^l t_j^2} \right).$$

In addition,  $\sigma_\epsilon^2$  can be estimated as

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{j=1}^l \hat{\sigma}_{\epsilon_i}^2, \quad (7)$$

where

$$\hat{\sigma}_{\epsilon_i}^2 = \frac{1}{l-1} \sum_{j=1}^l (y_{ij} - \hat{\beta}_i t_j)^2. \quad (8)$$

Now, we need to find an approximation of the unconditional distribution of  $\hat{\beta}_i$  and  $\log \hat{\beta}_i$ , in order to use them in the estimation of  $m$  and  $u$ . Therefore, by using Chebychev's inequality, it can be shown that  $\hat{\beta}_i$  converges in probability to  $\beta_i$  as

$$\begin{aligned} \Pr \left( \left( \hat{\beta}_i - \beta_i \right)^2 \geq \varepsilon \right) &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \hat{\beta}_i - \beta_i \right)^2 \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \mathbb{E} \left( \hat{\beta}_i - \beta_i \right)^2 | \beta_i \right) \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \text{var}(\hat{\beta}_i | \beta_i) \right) \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \frac{\sigma_\epsilon^2}{\sum_{j=1}^l t_j^2} \right). \end{aligned}$$

From the above inequality and the definition of convergence in probability, we can show that  $\hat{\beta}_i$  converges in probability to  $\beta_i$  when  $\sum_{j=1}^l t_j^2$  goes to infinity, i.e.

$$\hat{\beta}_i \xrightarrow{P} \beta_i \quad \text{as} \quad \sum_{j=1}^l t_j^2 \rightarrow \infty.$$

Moreover, since  $\log \hat{\beta}_i$  is a continuous function of  $\hat{\beta}_i$  we can also conclude that

$$\log \hat{\beta}_i \xrightarrow{P} \log \beta_i \quad \text{as} \quad \sum_{j=1}^l t_j^2 \rightarrow \infty.$$

According to the above result, it can be concluded that the asymptotic distribution of  $\log \hat{\beta}_i$  is the logistic distribution with location parameter  $m$  and scale parameter  $u$ , (i.e.  $\log \hat{\beta}_i \sim \text{Logistic}(m, u)$ ).

Now, in order to find the maximum likelihood estimator of  $m$  and  $u$ , let  $x_i = \log \hat{\beta}_i$ . Then

$$f(x_i) = \frac{e^{-(x_i-m)/u}}{u(1 + e^{-(x_i-m)/u})^2}, \quad x_i \in \mathbb{R} \quad (9)$$

and the log-likelihood function is

$$\log L(m, u) = \sum_{i=1}^n \left( -\frac{x_i - m}{u} - \log u - 2 \log (1 + e^{-(x_i-m)/u}) \right). \quad (10)$$

To find the maximum likelihood estimator (MLE)  $(\hat{m}, \hat{u})$ , we have to take the first partial derivatives of the log-likelihood with respect to  $m$  and  $u$ , and then set each derivative to zero and solve the resulting two equations simultaneously, i.e.

$$\frac{n}{2} = \sum_{i=1}^n \frac{e^{-(x_i-m)/u}}{1 + e^{-(x_i-m)/u}} \quad (11)$$

$$nu = \sum_{i=1}^n (x_i - m) \left( 1 - \frac{2}{1 + e^{(x_i-m)/u}} \right). \quad (12)$$

The estimators  $\hat{m}$  and  $\hat{u}$  can be obtained numerically by finding the roots of equations (11) and (12). Moreover, by substituting these two estimators into equation (5), an estimator of  $q_p$  is

$$\hat{q}_p = \exp \left( \log D - \hat{m} + \hat{u} \log \frac{p}{1-p} \right). \quad (13)$$

### 3.2 The Sampling Distribution of $(\hat{m}, \hat{u})$

Since  $\hat{m}$  and  $\hat{u}$  are the MLE's of  $m$  and  $u$ , we can use the asymptotic property of MLE's to find their sampling distribution, that is

$$\begin{pmatrix} \hat{m} \\ \hat{u} \end{pmatrix} \sim N \left( \begin{pmatrix} m \\ u \end{pmatrix}, \Sigma \right),$$

where

$$\Sigma = \begin{pmatrix} \text{var}(\hat{m}) & \text{cov}(\hat{m}, \hat{u}) \\ \text{cov}(\hat{m}, \hat{u}) & \text{var}(\hat{u}) \end{pmatrix}$$

denotes the variance-covariance matrix of  $\hat{m}$  and  $\hat{u}$ , and can be obtained by taking the inverse of the Fisher information matrix  $I^{-1}(m, u)$ .

The Fisher information matrix  $I(m, u)$  can be expressed as

$$I(m, u) = - \begin{pmatrix} E \left( \frac{\partial^2 \log L(m, u)}{\partial m^2} \right) & E \left( \frac{\partial^2 \log L(m, u)}{\partial m \partial u} \right) \\ E \left( \frac{\partial^2 \log L(m, u)}{\partial m \partial u} \right) & E \left( \frac{\partial^2 \log L(m, u)}{\partial u^2} \right) \end{pmatrix} .$$

Hence, the sampling distribution of  $\hat{m}$  and  $\hat{u}$  is given by

$$\begin{pmatrix} \hat{m} \\ \hat{u} \end{pmatrix} \sim N \left( \begin{pmatrix} m \\ u \end{pmatrix}, \begin{pmatrix} \frac{3u^2}{n} & 0 \\ 0 & \frac{3u^2}{n(1+u^2)} \end{pmatrix} \right) .$$

In a real situation the experiment is only conducted up to time  $t_l$ . Thus, the parameters  $(m, u)$  should be calibrated by using the conditional expectation technique (Yu and Tseng, 2004). The following equations describe  $(m_l, u_l)$  which denote the parameters after calibration (see Appendix)

$$m_l \approx m$$

and

$$u_l \approx \sqrt{u^2 + \frac{3}{\pi^2} \frac{\sigma^2 e^{-2m}}{\sum_{j=1}^l t_{ij}^2}} .$$

### 3.3 The Distribution of $\hat{q}_p$

We can use the calibrated parameters to find the asymptotic distribution of  $\log \hat{q}_p$ , which in turn will lead us to the distribution of  $\hat{q}_p$ .

Since  $\hat{m}$  and  $\hat{u}$  are jointly normal, then any linear combination of them has a normal distribution, too. Therefore,  $\log \hat{q}_p$  and its asymptotic distribution are expressed as

$$\log \hat{q}_p = \log D - \hat{m} + \hat{u} \log \frac{p}{1-p} \sim N(\mu, v) ,$$

where

$$\mu = \log D - m_l + u_l \log \frac{p}{1-p}$$

and

$$v = \frac{3u_l^2}{n} \left( 1 + \frac{1}{1+u_l^2} \left( \log \frac{p}{1-p} \right)^2 \right) .$$

Now, since  $\log \hat{q}_p \sim N(\mu, v)$  then  $\hat{q}_p$  follows a log-normal distribution with location parameter  $\mu$  and scale parameter  $\nu$  (i.e.  $\hat{q}_p \sim \text{Log-normal}(\mu, \nu)$ ).

It is easy to show that the mean and variance of  $\hat{q}_p$ , respectively, are

$$E(\hat{q}_p) = e^{\mu + \frac{\nu}{2}}$$

and

$$\text{var}(\hat{q}_p) = (e^\nu - 1)e^{2\mu + \nu} .$$

Thus, the mean squared error of the estimated 100 $p^{\text{th}}$  percentile  $\text{MSE}(\hat{q}_p(f, l, n))$  can be expressed as

$$\begin{aligned} \text{MSE}(\hat{q}_p(f, l, n)) &= \text{var}(\hat{q}_p) + E^2(\hat{q}_p - q_p) \\ &= (e^\nu - 1)e^{2\mu + \nu} + (e^{\mu + \frac{\nu}{2}} - e^\mu)^2 \\ &= e^{2\mu} (1 - 2e^{\frac{\nu}{2}} + e^{2\nu}) . \end{aligned}$$

### 3.4 The Cost Function $TC(f, l, n)$

The total cost of a degradation experiment consists of three main parts (Wu and Chang, 2002):

1. Cost of tested units. Let  $C_{it}$  denotes the cost of one unit. Then the total sample cost is  $nC_{it}$ .
2. Cost of inspections, which includes the cost of using inspection equipments and material. Let  $C_{mea}$  be the cost of one inspection on one unit, then the total inspection cost is  $nlC_{mea}$ .
3. Cost of operating the experiment, which consists of the salaries of operators and the cost of utility and depreciation of equipments. It depends on the termination time  $(l - 1)f$ . Let  $C_{op}$  be the cost of the operation in the interval between two inspections. Then  $(l - 1)fC_{op}$  is the total cost of the operation.

Thus, the total experimental cost can be obtained by combining the above subtotal costs in one equation as

$$TC(f, l, n) = nC_{it} + nlC_{mea} + (l - 1)fC_{op}.$$

According to the above results, we can rewrite the optimization problem (1) as

$$\begin{aligned} \text{minimize } e^{2\mu}(1 - 2e^{\frac{\nu}{2}} + e^{2\nu}) \quad \text{subject to } nC_{it} + nlC_{mea} + (l - 1)fC_{op} &\leq C_b, \\ f, l, n \in \mathbb{N}, \quad l \geq 2, \quad f \geq 1. \end{aligned}$$

But it is obvious that our optimization problem cannot be solved in a closed form easily, since both the objective function and the constraint are nonlinear functions of  $f$ ,  $l$ , and  $n$ . Therefore, in order to find the optimal solution for the problem, we will use the algorithm which was proposed by Wu and Chang (2002).

## 4 Simulation Study

A simulation study is performed to investigate and analyze our proposed method, Therefore, assume that  $(m, u, \sigma_\epsilon^2) = (2.0, 1.0, 2.5)$ , set  $p = 0.1$  and  $D = 100$ , and apply the algorithm which was proposed by Wu and Chang (2002), to get the optimal setting of the decision variables and their corresponding minimum  $MSE(\hat{q}_p(f, l, n))$ .

Changes in the values of  $C_b$ ,  $C_{op}$ ,  $C_{mea}$  and  $C_{it}$  affect the determination of the optimal degradation design. Thus, we examined the sensitivity of the decision variables  $(f, l, n)$  to the changes in cost elements. Results are represented in Table 1. From these results we can draw the following conclusions:

1. The sample size  $n$  is not affected by changes in  $C_{op}$  and  $C_{mea}$ , but it is highly affected by changes in  $C_{it}$  and  $C_b$ . In addition,  $n$  increases with  $C_b$  and decreases with  $C_{it}$ .
2. The number of inspections  $l$  is slightly affected by changes in  $C_{op}$  and highly affected by changes in the values of  $C_b$ ,  $C_{mea}$  and  $C_{it}$ . It is obvious that  $l$  decreases with both  $C_b$  and  $C_{mea}$ .

Table 1: Optimal values of  $f, l, n$  and termination time  $(l - 1)f$  under various values of the budget  $C_b$ , the operation cost  $C_{op}$ , the inspection cost  $C_{mea}$ , and the unit cost  $C_{it}$ .

$C_b$	$C_{op}$	$C_{mea}$	$C_{it}$	$f^*$	$l^*$	$n^*$	$(l^* - 1)f^*$	$MSE(\hat{q}_p(f^*, l^*, n^*))$
1000	0.5	0.06	50	5	14	19	65	2.98565
1200	0.5	0.06	50	8	10	23	72	2.12061
1400	0.5	0.06	50	16	6	27	80	1.62293
1600	0.5	0.06	50	10	8	31	70	1.30481
1800	0.5	0.06	50	9	8	35	63	1.08614
1000	0.5	0.06	50	5	14	19	65	2.98565
1000	0.7	0.06	50	3	16	19	45	2.98574
1000	0.9	0.06	50	4	11	19	40	2.98589
1000	1.1	0.06	50	3	12	19	33	2.98601
1000	1.3	0.06	50	2	14	19	26	2.98619
1000	0.5	0.06	50	5	14	19	65	2.98565
1000	0.5	0.08	50	9	9	19	72	2.98568
1000	0.5	0.10	50	12	7	19	72	2.98571
1000	0.5	0.12	50	9	8	19	63	2.98573
1000	0.5	0.14	50	18	5	19	72	2.98576
1000	0.5	0.06	50	5	14	19	65	2.98565
1000	0.5	0.06	60	4	14	16	52	4.16761
1000	0.5	0.06	70	3	9	14	24	5.51072
1000	0.5	0.06	80	5	13	12	60	7.79611
1000	0.5	0.06	90	8	3	11	16	9.63941

3. The inspection frequency  $f$  is sensitive to all cost elements. However, it nearly increases with  $C_b$  and  $C_{mea}$ , but decreases with  $C_{op}$ .
4. The termination time is affected by changes in all cost elements, and it decreases with  $C_{op}$  and  $C_{it}$ .
5. The value of the mean squared error is highly affected by the sample size.

## 5 Application to Real Data

One of the formally documented procedures in degradation analysis concerns the determination of the shelf lives of drugs (Chao, 1999). All drugs are specifically labeled to be used before a certain date. How are these expiration dates determined?

Developing a new drug involves performing a stability study to determine the drug shelf life (or lifetime). The expiration dating period or shelf life of a drug is defined as the length of time it takes for the drug’s potency to decrease to a particular level of its original potency. Potency refers to the amount or dose of a drug required to produce a given effect.

In the pharmaceutical industry, drug products are usually manufactured in different batches. The Food and Drug Administration (FDA) requires testing of at least three batches, preferably more, depending on economic considerations.

Therefore, assessing the lifetime distribution of a drug product and estimating its shelf life efficiently require an optimal design. Which in turn can be achieved by determining the optimal setting of the number of batches to be tested, the number of inspections, and the interval between inspections (or the inspection frequency), taking into consideration that the total cost of the test does not exceed a particular budget.

## 5.1 Drug Potency Data

Consider a typical set of data, reported by Chow and Shao (1991) that obtained from a stability study of a certain drug. There is a random sample of 24 batches; measurements of their potency are taken at 0, 12, 24, and 36 months. The main purpose of analyzing this data is to determine the shelf life of this drug, defining its shelf life (or lifetime) as the length of time it takes for this drug's potency to decrease to 95% of its original stated potency. Consequently, the critical degradation level is 95.

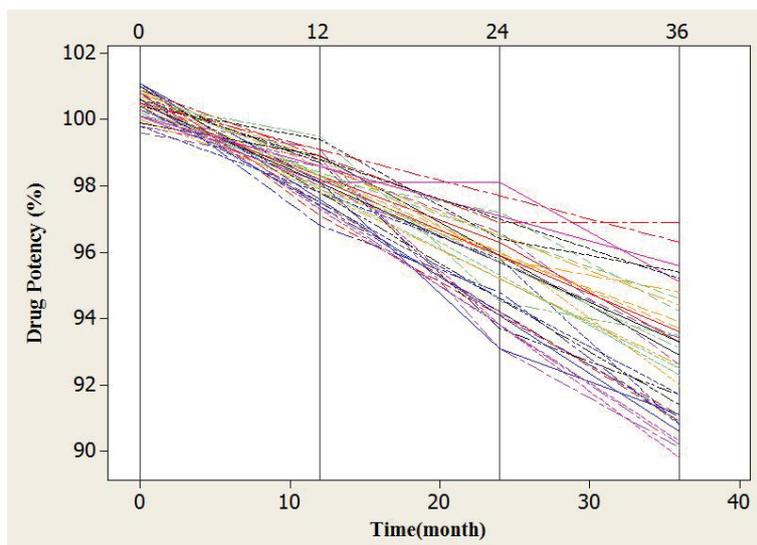


Figure 1: Plot of drug potency degradation data over time  $t$  in months.

The drug's potency degradation data is plotted in Figure 1, which shows that the degradation curves are nearly linear, thus, we can use a linear degradation model to describe the degradation observation. Clearly, the degradation process in the above data is a decreasing process and the degradation rate has a negative sign. Therefore, in order to make this data applicable to our model (2), we should rescale the data by replacing the drug's potency degradation data and use the reduction in drug potency data instead.

Let  $w_{ij}$  denotes the reduction of drug potency of the  $i^{\text{th}}$  batch at time  $t_j$  (i.e.  $w_{ij} = 100 - y_{ij}$ ), where  $y_{ij}$  is the original drug potency of the  $i^{\text{th}}$  batch at time  $t_j$ . Consider the transformed data as plotted in Figure 2, we can use model (2) to describe the reduction of drug potency as

$$w_{ij} = \beta_i t_j + \epsilon_{ij},$$

where  $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$  and  $\beta_i$  follows the log-logistic distribution.

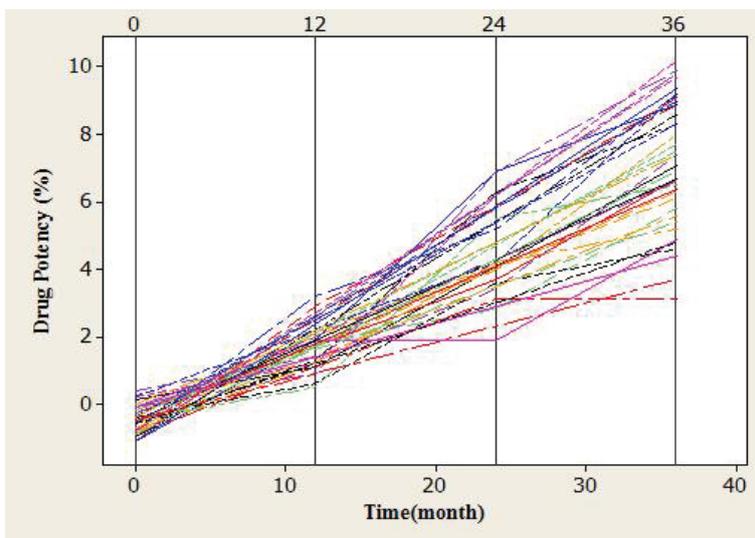


Figure 2: Plot of reduction of drug potency over time  $t$  in months.

### 5.2 Estimation of $\sigma_\epsilon^2$ , $m$ , and $u$

Based on the observations  $(t_j, w_{ij})$ ,  $j = 1, \dots, 4$ , and by using equations (6) and (7), the least squares estimators  $\hat{\beta}_i$  and  $\hat{\sigma}_{\epsilon_i}^2$  can be obtained, respectively. Figure 3 shows the logistic probability plot for  $\log \hat{\beta}_i$ ,  $i = 1, \dots, 24$ . The linear pattern of the plot indicates that it is reasonable to assume that  $\beta_i$  follows the log-logistic distribution.

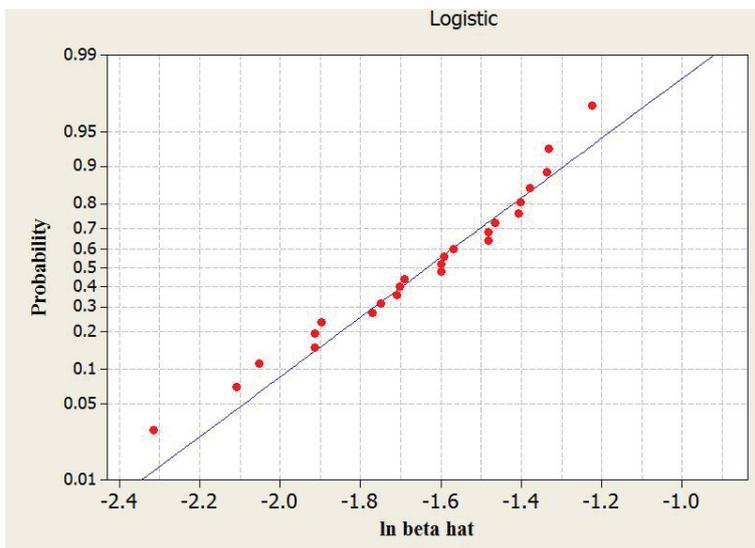


Figure 3: The logistic probability plot for  $\log \hat{\beta}_i$ ,  $i = 1, \dots, 24$ .

Now, from equations (7), (11), and (12),  $\sigma_\epsilon^2$  and the parameters  $m$  and  $u$  can be estimated as

$$(\hat{\sigma}_\epsilon^2, \hat{m}, \hat{u}) = (0.82004, -1.63384, 0.154635).$$

### 5.3 Optimal Test Plan Based on the Reduction of Drug's Potency Data

As we mentioned before that the shelf life of a drug is the time it takes for the drug's potency to decrease to 95% of its original stated potency, i.e. it is the time when the amount of reduction in potency is 5. Thus, in this case our critical degradation level  $D = 5$ .

Moreover, assume that the cost elements needed for conducting this test are as follows:  $C_{it} = 30$ ,  $C_{op} = 18$ ,  $C_{mea} = 8$ , and  $C_b = 1000$ . Finally, take  $p = 0.1$  and let the unit of time be measured by months. The optimal setting of the decision variables and their corresponding minimum  $MSE(\hat{q}_p(f, l, n))$  are listed in Table 2.

Table 2: Optimal setting of the decision variables for the drug's potency data.

$C_b$	$C_{op}$	$C_{mea}$	$C_{it}$	$f^*$	$l^*$	$n^*$	$(l^* - 1)f^*$	$MSE(\hat{q}_p(f^*, l^*, n^*))$
1000	18	8	30	8	5	6	32	98.0102

That is, the optimal inspection interval  $f^* = 8$  months, the optimal number of batches  $n^* = 6$ , the optimal number of inspection  $l^* = 5$ , and the optimal termination time of the test is 32 months. In other words, in order to estimate the 10<sup>th</sup> percentile of the drug's lifetime distribution, or to estimate the time at which 10% of the drug batches will lose 5% of their original potency and considered as expired, we need to test 6 batches during 32 months by taking measurements of their drug potency every 8 months.

## 6 Concluding Remarks

In this paper, we obtained the optimal design of a degradation experiment where the degradation rate follows the log-logistic distribution. Particularly, we derived the optimal setting of the decision variables: the sample size, the inspection frequency, and the termination time, by using the criterion of minimizing the mean squared error of the estimated 100<sup>th</sup> percentile of the products lifetime subject to the constraint that the total cost does not exceed a predetermined budget. Some concluding remarks are given as:

1. Selecting the optimal combination of the sample size, the inspection frequency, and the number of inspections with the minimal cost are the key points for optimal degradation experiment design.
2. Among all decision variables  $f$ ,  $l$ , and  $n$ , the sample size  $n$  has the most important influence on the value of the mean squared error. Therefore, if we wish to reduce the mean squared error, we should increase the number of test units. In other words, if we want to increase the precision we have to raise the budget of the experiment or we can reduce the cost of tested units.
3. The budget of the experiment has a high influence on most of the decision variables. Thus, if we raise the budget of the experiments the number of units will increase, but the number of inspections and the mean squared error will decrease.

## References

- Al-Haj Ebrahim, M., Alodat, M., and Arman, A. (2009a). Estimating the time-to-failure distribution of a linear degradation model using a Bayesian approach. *Applied Mathematical Sciences*, 3, 27-42.
- Al-Haj Ebrahim, M., Eidous, O., and Kamil, G. (2009b). Estimating percentiles of time-to-failure distribution obtained from a linear degradation model using kernel density method. *Communications in Statistics – Simulation and Computation*, 38, 1811-1822.
- Chao, M. T. (1999). Degradation analysis and related topics: some thoughts and review. *Proceedings of the National Science Council*, 23, 555-566.
- Chiodo, E., and Mazzanti, G. (2004). The log-logistic model for reliability characterization of power system components subjected to random stress. *Speedam, Capri/Italy*, 239-244.
- Chow, S. C., and Shao, J. (1991). Estimating drug shelf-life with random effects. *Biometrics*, 47, 1071-1079.
- Lu, C. J., and Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35, 161-174.
- Meeker, W. Q., and Escobar, L. A. (1998). *Statistical methods for reliability data*. New York: John Wiley.
- Wu, S. J., and Chang, C. T. (2002). Optimal design of degradation tests in presence of cost constraint. *Reliability Engineering & System Safety*, 76, 109-115.
- Yu, H. F. (2002a). Designing an accelerated degradation experiment by optimizing the interval estimation of the mean-time-to-failure. *Journal of Chinese Institute of Industrial Engineers*, 19, 23-33.
- Yu, H. F. (2002b). Designing an accelerated degradation experiment with reciprocal Weibull degradation rate, journal=Journal of Statistical Planning and Inference, volume=136, pages=282-297.
- Yu, H. F. (2003). Designing an accelerated degradation experiment by optimizing the estimation of the percentile. *Quality and Reliability Engineering International*, 19, 197-214.
- Yu, H. F., and Tseng, S. T. (1999). Designing a degradation experiment. *Naval Research Logistics*, 46, 689-706.
- Yu, H. F., and Tseng, S. T. (2004). Designing a degradation experiment with a reciprocal Weibull degradation rate. *Quality Technology & Quantitative Management*, 1, 47-63.

## Appendix

Assume that  $\log \hat{\beta}_i \sim \text{Logistic}(m_l, u_l)$ . Then we have

$$E(\log \hat{\beta}_i) = m_l \quad (14)$$

and

$$\text{var}(\log \hat{\beta}_i) = \frac{\pi^2}{3} u_l^2. \quad (15)$$

We use the following conditional expectation rules to find the unconditional mean and variance of  $\log \hat{\beta}_i$ , i.e.

$$E(\log \hat{\beta}_i) = E(E(\log \hat{\beta}_i | \beta_i)) \approx E(\log \beta_i) = m \quad (16)$$

and

$$\begin{aligned} \text{var}(\log \hat{\beta}_i) &= \text{var}(E(\log \hat{\beta}_i | \beta_i)) + (\text{var}(\log \hat{\beta}_i | \beta_i)) \\ &\approx \text{var}(\log \beta_i) + E\left(\frac{\sigma_\epsilon^2}{\beta_i^2 \sum_{j=1}^l t_j^2}\right) \\ &= \frac{\pi^2}{3} u^2 + \frac{\sigma_\epsilon^2}{\sum_{j=1}^l t_j^2} E\left(\frac{1}{\beta_i^2}\right) \\ &\approx \frac{\pi^2}{3} u^2 + \frac{\sigma_\epsilon^2 e^{-2m}}{\sum_{j=1}^l t_j^2}. \end{aligned} \quad (17)$$

Equate equations (14) with (16) and (15) with (17) to get

$$m_l \approx m$$

and

$$\begin{aligned} \frac{\pi^2}{3} u_l^2 &\approx \frac{\pi^2}{3} u^2 + \frac{\sigma_\epsilon^2 e^{-2m}}{\sum_{j=1}^l t_j^2} \\ u_l &\approx \sqrt{u^2 + \frac{3 \sigma_\epsilon^2 e^{-2m}}{\pi^2 \sum_{j=1}^l t_j^2}}. \end{aligned}$$

Authors' address:

Rana Azmi Dandis, Mohammed Al-Haj Ebrahim

Department of Statistics

Yarmouk University

Irbid

Jordan

E-mails: rana\_dandis@yahoo.com, m\_hassanb@hotmail.com