Chapter 3
Statistical Inference

3.1 Introduction

Inference about the values of parameters involved in statistical distributions is known as estimation. The engineer might be interested in estimating the mean life of an electronic device based on the failure times of a random sample placed on lifef test. An interesting question that commonly would ask from many practitioners is: how close this estimator would be the true value of the parameter being estimated from a known distribution.

An estimator is a procedure which provides an estimate of a population parameter from a random sample. In other word, an estimator is a statistic and hence a random variable, since it depends strictly on the sample. An estimator is called a point estimator if it provides a single value as an estimate from a random sample.

In this chapter, it is assumed that the population distribution by type is known, but the distribution parameters are unknown and they have to be estimated by using collected failure data. This chapter is devoted to the theory of estimation and discusses several common estimation techniques such as maximum likelihood, method of moments, least squared, and Bayesian methods. We also discuss the confidence interval estimates, tolerance limit estimates, sequential sampling and criteria for model selection.

3.2 Statistical Inference

Statistical inference is the process of drawing conclusions about unknown characteristics of a population from which data were taken. Techniques used in this process include paramter estimation, confidence intervals, hypothesis testing, goodness of fit tests, and sequential testing. As we know, any distribution function often involves some parameters. The problem of point estimation is that of estimating the parameters of a population. For example, parameter λ from an exponential distribution; μ and

H. Pham, *Statistical Reliability Engineering*, Springer Series in Reliability Engineering,
https://doi.org/10.1007/978-3-030-76904-8_3

σ^2 from a normal distribution; or n and p from the binomial. For simplicity, denote these parameters by the notation $\boldsymbol{\theta}$ which may be a vector consisting of several distribution parameters. Given a real life application and the observed data, the common statistical problem will consist of how to determine the unknown distribution F or pdf f. And if F is assumed to be known, then how can one determine the unknown distribution parameters $\boldsymbol{\theta}$. If a statistic $Y = h(X_1, X_2,\ldots, X_n)$ is taken as an estimate of the parameters $\boldsymbol{\theta}$, the first point to recognize is that in any particular sample the observed value of $h(X_1, X_2,\ldots, X_n)$ may differ from $\boldsymbol{\theta}$. The performance of $h(X_1, X_2,\ldots, X_n)$ as an estimate of $\boldsymbol{\theta}$ is to be judged in relation to the sampling distribution of $h(X_1, X_2,\ldots, X_n)$. For example, assume n independent samples from the exponential density $f(x; \lambda) = \lambda e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$, then the joint pdf or sample density (for short) is given by

$$f(x_1, \lambda) \cdot f(x_1, \lambda) \ldots f(x_1, \lambda) = \lambda^n e^{-\lambda \sum_{i-1}^{n} x_i} \tag{3.1}$$

The problem here is to find a "good" point estimate of λ which is denoted by $\hat{\lambda}$. In other words, we shall find a function $h(X_1, X_2,\ldots,X_n)$ such that, if $x_1, x_2, \ldots, x_n$ are the observed experimental values of $X_1, X_2, \ldots, X_n$, then the value $h(x_1, x_2, \ldots, x_n)$ will be a good point estimate of λ. By "good' we mean the following properties shall be implied:

- Unbiasedness
- Consistency
- Efficiency (i.e., minimum variance)
- Sufficiency
- Asymptotic efficiency.

In other words, if $\hat{\lambda}$ is a good point estimate of λ, then one can select the function $h(X_1, X_2,\ldots,X_n)$ such that $h(X_1, X_2,\ldots,X_n)$ is not only an unbiased estimator of λ but also the variance of $h(X_1, X_2,\ldots,X_n)$ is a minimum. It should be noted that the estimator $\hat{\lambda}$ is a random variable where λ is a number.

The random variable $h(X_1, X_2,\ldots,X_n)$ is called a statistic and it represents the estimator of the unknown parameters $\boldsymbol{\theta}$. We will now present the following definitions.

Definition 3.1 For a given positive integer n, the statistic $Y = h(X_1, X_2,\ldots, X_n)$ is called an *unbiased* estimator of the parameter $\boldsymbol{\theta}$ if the expectation of Y is equal to a parameter $\boldsymbol{\theta}$, that is,

$$E(Y) = \theta \tag{3.2}$$

In other words, an estimator of $\boldsymbol{\theta}$ is called unbiased if its expected value is equal to the population value of the quantity it estimates.

An estimator Y is a best unbiased estimator of parameter $\boldsymbol{\theta}$ if $E(Y) = \theta$ for all $\theta \in \Theta$ and $\text{var}(Y) \leq \text{var}\left(\tilde{Y}\right)$ for any other unbiased estimator $\tilde{Y}$ such that.

$E\left(\tilde{Y}\right) = \theta$. If an estimator is not unbiased we can define its *bias* by

$$b(\theta) = E(Y) - \theta.$$

It sometimes happens that $b(\theta)$ depends on the number of observations and approaches to zero as n increases. In this case, Y is said to be *asymptotically unbiased*. Note that the expectation represents a long-run average, bias is seen as an average deviation of the estimator from the true value. When considering bias, care needs to be taken with functions of parameters. In general, if $E(Y) = \theta$ and g(θ) is some function of a parameter,

$$E[g(Y)] \neq g(\theta).$$

For the exponential distribution for example, assume X_1,X_2, …, X_n be a random sample from the exponential distribution with pdf

$$f(x;\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \quad x > 0, \mu > 0$$

Let $Y = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$. Using the maximum likelihood estimation (MLE) method (see Sect. 3.3.2), we can show that Y is an unbiased estimator of μ. That is, $E(Y) = \mu$.

Let $g(Y) = \frac{n}{\sum_{i=1}^{n} x_i}$. It should be noted that $E[g(Y)] \neq \frac{1}{\mu}$.

Definition 3.2 The statistic Y is said to be a *consistent* estimator of the parameter $\boldsymbol{\theta}$ if Y converges stochastically to a parameter $\boldsymbol{\theta}$ as n approaches infinity. If $\in$ is an arbitrarily small positive number when Y is consistent, then.

$$\lim_{n\to\infty} P(|Y - \theta| \leq \in) = 1 \tag{3.3}$$

In other words, we would like our estimators tend towards the true value when the sample size n closes to infinity or to a very large number. It can be shown that this will happen if the estimator Y is either unbiased or asymptotically unbiased and also has a variance that tends to zero as n increases.

Example 3.1 Let X be a life time with mean μ which is unknown. One can easily show that the statistic.

$$h(X_1, X_2, \ldots, X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

is unbiased and a consistent of μ.

Definition 3.3 If the limiting distribution of.

$$\frac{\sqrt{n}(Y_n - \theta)}{a} \sim N(0, 1)$$

where Y_n is an estimator of $\boldsymbol{\theta}$ based on a sample of size n, the *asymptotic efficiency* of Y_n is defined as

$$AE = \frac{1}{a^2 I(\theta)} \text{ where } I(\theta) = -E\left(\frac{\partial^2 \ln f}{\partial \theta^2}\right) \tag{3.4}$$

and f is the probability density function of the parent population. If $AE = 1$, the sequence of estimators is said to be asymptotically efficient or best asymptotically normal (BAN).

Definition 3.4 The statistic Y is said to be *sufficient* for $\boldsymbol{\theta}$ if the conditional distribution of X, given $Y = y$, is independent of $\boldsymbol{\theta}$.

Definition 3.5 The statistic Y will be called the *minimum variance unbiased estimator* of the parameter $\boldsymbol{\theta}$ if Y is unbiased and the variance of Y is less than or equal to the variance of every other unbiased estimator of $\boldsymbol{\theta}$. An estimator that has the property of minimum variance in large samples is said to be efficient.

In other words, among unbiased estimates of $\boldsymbol{\theta}$, the one that has smallest variance is preferred. For the variance of any unbiased estimator Y of $\boldsymbol{\theta}$, we have the lower limit

$$Var(Y) \geq \frac{1}{n\, I(\theta)} \tag{3.5}$$

where $I(\theta)$, defined by Eq. (3.4), is called the amount of information per observation (also see Sect. 3.4).

Definition 3.5 is useful in finding a lower bound on the variance of all unbiased estimators. In fact, a *minimum variance unbiased estimator* is one whose variance is least for all possible choices of unbiased estimators. Let X be a random variable of life time with mean μ which is unknown. Consider the following two statistics

$$h_1(X_1, X_2, \ldots, X_n) = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3$$

and

$$h_2(X_1, X_2, \ldots, X_n) = \frac{1}{3}(X_1 + X_2 + X_3)$$

We can easily show that both h_1 and h_2 give unbiased estimates of μ. However, based on the unbiased estimates, we cannot tell which one is better, but we can tell

which estimator is better by comparing the variances of both h_1 and h_2 estimators. It is easy to see that var(h_1) > var(h_2), then h_2 is a better estimator h_1. Mainly, the smaller the variance the better!

The mean squared error is also another criterion for comparing estimators by combining the variance and the biased parts.

Definition 3.6 The *mean squared error* (MSE) is defined as the expected value of the square of the deviation of the estimate from the parameter being estimated, and is equal to the variance of the estimate plus the square of the bias. That is,

$$MSE = E\left(\hat{\theta} - \theta\right)^2 = \text{var}\left(\hat{\theta}\right) + E\left[E\left(\hat{\theta}\right) - \theta\right]^2 = \text{var}\left(\hat{\theta}\right) + (bias)^2.$$

Obviously, a small value of MSE is a desirable feature for an estimator. For unbiased estimators case, seeking a small MSE is identical to seeking a small variance since the second term on the right hand side is equal to zero.

Example 3.2
Assume $\hat{\theta}_1$ and $\hat{\theta}_2$ are the two estimators of parameter θ. Suppose that $E\left(\hat{\theta}_1\right) = 0.9\theta, E\left(\hat{\theta}_2\right) = \theta$, var($\hat{\theta}_1$) = 2, and var($\hat{\theta}_2$) = 3. Which estimator would you prefer?

Solution: We have

$$MSE \text{ of } \theta_1 = \text{var}\left(\hat{\theta}\right) + (bias)^2 = 2 + 0.01\,\theta^2.$$
$$MSE \text{ of } \theta_2 = \text{var}\left(\hat{\theta}\right) + (bias)^2 = 3 + 0 = 3.$$

Thus, if $|\theta| < 10$ then $\hat{\theta}_1$ would be prefered. If $|\theta| > 10$ then $\hat{\theta}_2$ is prefered.

We will later discuss how to establish a lower bound on the variance using an inequality known as the Cramér-Rao inequality. We now discuss some basic methods of parameter estimation.

3.3 Parameter Estimation

Once a distribution function is specified with its parameters, and data have been collected, one is in a position to evaluate its goodness of fit, that is, how well it fits the observed data. The goodness of fit is assessed by finding parameter values of a model that best fits the data—a procedure called parameter estimation. There are two general methods of parameter estimation. They are the methods of moments and maximum likelihood estimate (MLE).

3.3.1 The Method of Moments

The unknown distribution parameters usually can be estimated by their respective moments, such as means, variances etc. In general, we can choose the simplest moments to equate.

The expected value of a function $g(X)$ under the distribution $F(x)$ is given by

$$E(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x) f(x) dx & \text{if X is continuous} \\ \sum_{i=0}^{n} g(x_i) p(x_i) & \text{if X is discrete} \end{cases} \tag{3.6}$$

The kth moment of $F(x)$ is defined as follows:

$$\mu_k = E(X^k) \quad \text{for k = 1, 2, ...} \tag{3.7}$$

Obviously, when $k = 1$ it is the expected value of X, that is $\mu_1 = E(X)$. When X only assumes positive values and has a continuous pdf $f(x)$ and cdf $F(x)$, then

$$\begin{aligned} E(X) &= \int_0^{\infty} x f(x) dx = \int_0^{\infty} \left(\int_0^{x} dy \right) f(x) dx \\ &= \int_0^{\infty} \left(\int_y^{\infty} f(x) dx \right) dy = \int_0^{\infty} (1 - F(y)) dy. \end{aligned} \tag{3.8}$$

The kth central moments around the mean, μ_k^c, is defined as follows:

$$\mu_k^c = E[(X - \mu_1)^k] \text{ for } k = 1, 2, \ldots \tag{3.9}$$

The 1st central moment is *0*. The second central moment is exactly the variance of $F(x)$. That is,

$$\mu_2^c = E[(X - \mu_1)^2]. \tag{3.10}$$

Example 3.3
Suppose that $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0, \lambda > 0$. Then $F(x) = 1 - e^{-\lambda x}$ and the population mean (or the first moment) is.

$$\mu_1 = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$$

The sample mean is $\bar{x}$ so the estimator $\hat{\lambda}$ of λ that makes these two the same is

$$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Note that when X is discrete, assuming the values $\{1,2,\ldots\}$, then we obtain

$$E(X) = 1 + \sum_{i=1}^{\infty} [1 - F(i)].$$

Example 3.4
For a normal distribution with two unknown parameters pdf, i.e. $N(\mu, \sigma^2)$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

Then the first and second moments about the mean of the sample are

$$\mu_1 = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and

$$\mu_2^c = E[(X - \mu_1)^2] = \frac{\sum_{i=1}^{n} (\mathrm{x_i} - \bar{x})^2}{n}.$$

For the population of corresponding moments are

$$\mu_1 = \mu \quad \text{and} \quad \mu_2^c = \sigma^2.$$

So the method of moments estimators are

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}.$$

3.3.2 Maximum Likelihood Estimation Method

The method of maximum likelihood estimation (MLE) is one of the most useful techniques for deriving point estimators. As a lead-in to this method, a simple example

will be considered. The assumption that the sample is representative of the population will be exercised both in the example and later discussions.

Example 3.5
Consider a sequence of 25 Bernoulli trials (binomial situation) where each trial results in either success or failure. From the 25 trials, 6 failures and 19 successes result. Let p be the probability of success, and 1–p the probability of failure. Find the estimator of p, $\hat{p}$, which maximizes that particular outcome.

Solution: The sample density function can be written as

$$g(19) = \binom{25}{19} p^{19}(1-p)^6.$$

The maximum of $g(19)$ occurs when

$$p = \hat{p} = \frac{19}{25}$$

so that

$$g\left(19|p = \frac{19}{25}\right) \geq g\left(19|p \neq \frac{19}{25}\right)$$

Now g(19) is the probability or "likelihood" of 6 failures in a sequence of 25 trials. Select $p = \hat{p} = \frac{19}{25}$ as the probability or likelihood maximum value and, hence, $\hat{p}$ is referred to as the maximum likelihood estimate. The reason for maximizing $g(19)$ is that the sample contained six failures, and hence, if it is representative of the population, it is desired to find an estimate which maximizes this sample result. Just as $g(19)$ was a particular sample estimate, in general, one deals with a sample density:

$$f(x_1, x_2, \ldots, x_n) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta) \tag{3.11}$$

where x_1, x_2, …, x_n are random, independent observation from a population with density function $f(x)$. For the general case, it is desired to find an estimate or estimates,$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ (if such exist) where

$$f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_m) > f(x_1, x_2, \ldots, x_n; \theta\prime_1, \theta\prime_2, \ldots, \theta\prime_m) \tag{3.12}$$

Notation $\theta'_1, \theta'_2, \ldots, \theta'_n$ refers to any other estimates different than $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$.

Once data have been collected and the likelihood function of a model given the data is determined, one is in a position to make statistical inferences about the population, that is, the probability distribution that underlies the data. We are interested in finding the parameter value that corresponds to the desired probability distribution. The principle of maximum likelihood estimation (MLE), originally developed by R. A.

Fisher in the 1920s, states that the desired probability distribution is the one that makes the observed data "most likely," which means that one must seek the value of the parameter vector that maximizes the likelihood function.

Let us now discuss the method of MLE. Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution having pdf $f(x; \theta)$. This distribution has a vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$'of unknown parameters associated with it, where *m is* the number of unknown parameters. Assuming that the random variables are independent, then the likelihood function, L(X; θ), is the product of the probability density function evaluated at each sample point:

$$L(X, \theta) = \prod_{i=1}^{n} f(X_i; \theta) \tag{3.13}$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The maximum likelihood estimator $\hat{\theta}$ is found by maximizing $L(\boldsymbol{X}; \theta)$ with respect to θ. In practice, it is often easier to maximize $\ln[L(\boldsymbol{X};\theta)]$ to find the vector of MLEs, which is valid because the logarithm function is monotonic. In other words, the maximum of $L(\boldsymbol{X}; \theta)$ will occur at the same value of θ as that for $\ln[L(\boldsymbol{X}; \theta)]$. The logarithm turns the multiplication of terms like $f(\boldsymbol{X_i}; \theta)$ into the addition of $\ln f(\boldsymbol{X_i}; \theta)$. Thus, the log likelihood function, denoted as $\ln L(\theta)$, is given by

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(X_i; \theta) \tag{3.14}$$

and is asymptotically normally distributed since it consists of the sum of n independent variables and the implication of the central limit theorem. Since L(X; θ) is a joint probability density function for $X_1, X_2, \ldots, X_n$, it must integrate equal to 1, that is,

$$\int_0^\infty \int_0^\infty \cdots \int_0^\infty L(X; \theta) dX = 1$$

Assuming that the likelihood is continuous, the partial derivative of the left-hand side with respect to one of the parameters, θ_i, yields

$$\frac{\partial}{\partial \theta_i} \int_0^\infty \int_0^\infty \cdots \int_0^\infty L(X; \theta) dX = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \frac{\partial}{\partial \theta_i} L(X; \theta) dX$$

$$= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \frac{\partial \ln L(X; \theta)}{\partial \theta_i} L(X; \theta) dX$$

$$= E\left[\frac{\partial \ln L(X;\theta)}{\partial \theta_i}\right]$$
$$= E[U_i(\theta)] \text{ for } i = 1, 2, \ldots, m \tag{3.15}$$

where $\boldsymbol{U}(\theta) = (U_1(\theta), U_2(\theta), \ldots U_n(\theta))$' is often called the score vector and the vector $\boldsymbol{U}(\theta)$ has components

$$U_i(\theta) = \frac{\partial[\ln L(X;\theta)]}{\partial \theta_i} \text{ for } i = 1, 2, \ldots, m \tag{3.16}$$

which, when equated to zero and solved, yields the MLE vector $\boldsymbol{\theta}$. In other words, the MLE vector $\boldsymbol{\theta}$ satisfy

$$\frac{\partial[\ln L(X;\theta)]}{\partial \theta_i} = 0 \text{ for } i = 1, 2, \ldots, m$$

This means that if more than one parameter is to be estimated, the partial derivatives with respect to each parameter are then set equal to zero and the resulting differential equations are solved for the estimates. It is worth to note that one needs to make sure whether the solution is actually a maximum and not a minimum or a point of inflection. Often, the maximum likehood estimators are biased.

Suppose that we can obtain a non-trivial function of $X_1, X_2, \ldots, X_n$, say $h(X_1, X_2, \ldots, X_n)$, such that, when θ is replaced by $h(X_1, X_2, \ldots, X_n)$, the likelihood function L will achieve a maximum. In other words,

$$L(X, h(X)) \geq L(X, \theta)$$

for every θ. The statistic $h(X_1, X_2, \ldots, X_n)$ is called a maximum likelihood estimator of θ and will be denoted as

$$\hat{\theta} = h(x_1, x_2, \ldots, x_n)$$

The observed value of $\hat{\theta}$ is called the MLE of θ. In other words, the MLE of θ is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ln L(\theta) \tag{3.17}$$

where Θ is the parameter space. Based on the asymptotic maximum likelihood theory, the MLE $\hat{\theta}$ is consistent and asymtotically efficient with limiting distribution

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \rightarrow N\left(0,\ I^{-1}(\theta_0)\right)$$

where θ_0 is the true parameter value and I is the Fisher information matrix. The asymptotic variance matrix $I^{-1}(\theta_0)$ can be estimated consistently by an empirical variance matrix of the influence functions evaluated at $\hat{\theta}$.

In general, the mechanics for obtaining the MLE can be obtained as follows:

Step 1. Find the joint density function $L(X, \theta)$.

Step 2. Take the natural log of the join density, ln $L(\theta)$.

Step 3. Take the partial derivatives of ln $L(\theta)$ (or $L(\theta)$) with respect to each parameter.

Step 4. Set partial derivatives to "zero".

Step 5. Solve for parameter(s).

Example 3.6

A sample of size n is drawn, without replacement, from a population of size N composed of k individuals of type 1 and (N-k) individuals of type 2. Assume that a population size N is unknown. The number X of individuals of type 1 in the sample is a hypergeometric random variable with pdf.

$$P[X = x] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \ldots, n \tag{3.18}$$

Obtain the MLE of N when k and n are known.

Solution: Let

$$P(x, N) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

Then

$$\frac{P(x, N)}{P(x, N-1)} = \frac{\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}}{\dfrac{\binom{k}{x}\binom{N-k-1}{n-x}}{\binom{N-1}{n}}}.$$

After simplifications, we obtain

$$\frac{P(x, N)}{P(x, N-1)} = \frac{N^2 - kN - nN + kn}{N^2 - kN - nN + xN}.$$

Note that P(x, N) is greater than, equal to, or less than P(x, $N-1$) according to kn is greater than, equal to, or less than xN, or equivalently, as N is less than, equal to, or greater than kn/x. We can now consider the following two cases.

Case 1: when $\frac{kn}{x}$ is not an integer.

The sequence $\{P(x, N), N = 1, 2, \ldots\}$ is increasing when $N < \frac{kn}{x}$ and is decreasing when $N > \frac{kn}{x}$. Thus, the maximum value of P(x, N) occurs when $N = \left\lfloor \frac{kn}{x} \right\rfloor$ which is the largest integer less than $\frac{kn}{x}$.

Case 2: when $\frac{kn}{x}$ is an integer.

In this case the maximum value of P(x, N) occurs when both P(x,N) and P(x, $N-1$) are equal. This implies that

$$\frac{kn}{x} - 1 = \left\lfloor \frac{kn}{x} \right\rfloor$$

as an estimate of the population. Therefore, the MLE of N is $\hat{N} = \left\lfloor \frac{kn}{x} \right\rfloor$ where $\lfloor y \rfloor$ denotes the greatest integer less than y.

Example 3.7

Let $X_1, X_2, \ldots, X_n$ be a random sample from the exponential distribu- tion with pdf.

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

The joint pdf of $X_1, X_2, \ldots, X_n$, is given by

$$L(X, \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

and

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

The function ln$L$ can be maximized by setting the first derivative of ln$L$, with respect to λ, equal to zero and solving the resulting equation for λ. Therefore,

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

This implies that

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

The observed value of $\hat{\lambda}$ is the MLE of λ.

Example 3.8
Let $X_1, X_2, \ldots, X_n$, denote a random sample from the normal distribution $N(\mu, \sigma^2)$ with unknown mean μ and known σ^2 where pdf is.

$$f(x; \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty, \ \mu \in (-\infty, \infty). \tag{3.19}$$

Find the maximum likelihood estimator of μ.
Solution: The likelihood function of a sample of size n is

$$L\left(\underset{\sim}{x}, \mu\right) = \prod_{i=1}^{n} f(x_i; \mu) = \prod_{i=1}^{n} \frac{1}{\left(\sigma\sqrt{2\pi}\right)} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$
$$= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

The log of likelihood function is

$$\ln L = -n\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Thus we have

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0.$$

Solving the above equation, we obtain

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

Thus, the MLE of μ is $\bar{x}$.

Example 3.9
In an exponential censored case, the non-conditional joint pdf that r items have failed is given by.

$$f(x_1, x_2, \ldots, x_r) = \lambda^r e^{-\lambda\sum_{i=1}^{r} x_i} \quad (r \text{ failed items}) \tag{3.20}$$

and the probability distribution that $(n-r)$ items will survive is

$$P(X_{r+1} > t_1, X_{r+2} > t_2, \ldots, X_n > t_{n-r}) = e^{-\lambda \sum_{j=1}^{n-r} t_j}$$

Thus, the joint density function is

$$\begin{aligned} L(X, \lambda) &= f(x_1, x_2, \ldots, x_r)\ P(X_{r+1} > t_1, \ldots, X_n > t_{n-r}) \\ &= \frac{n!}{(n-r)!} \lambda^r e^{-\lambda(\sum_{i=1}^{r} x_i + \sum_{j=1}^{n-r} t_j)} \end{aligned}$$

Let

$$T = \sum_{i=1}^{r} x_i + \sum_{j=1}^{n-r} t_j \tag{3.21}$$

then

$$\ln L = \ln\left(\frac{n!}{(n-r)!}\right) + r \ln \lambda - \lambda T$$

and

$$\frac{\partial \ln L}{\partial \lambda} = \frac{r}{\lambda} - T = 0.$$

Hence,

$$\hat{\lambda} = \frac{r}{T}. \tag{3.22}$$

Note that with the exponential, regardless of the censoring type or lack of censoring, the MLE of λ is the number of failures divided by the total operating time.

Example 3.10
Let $X_1, X_2, \ldots, X_n$ represent a random sample from the distribution with pdf.

$$f(x; \theta) = e^{-(x-\theta)} \quad \text{for } \theta \le x \le \infty \text{ and } -\infty < \theta < \infty. \tag{3.23}$$

The likelihood function is given by

$$L(\theta; X) = \prod_{i=1}^{n} f(x_i; \theta) \text{ for } \theta \le x_i \le \infty \text{ all } i$$

$$= \prod_{i=1}^{n} e^{-(x_i-\theta)} = e^{-\sum_{i=1}^{n} x_i + n\theta}.$$

For fixed values of $x_1, x_2, \ldots, x_n$, we wish to find that value of θ which maximizes $L(\theta; X)$. Here we cannot use the techniques of calculus to maximize $L(\theta; X)$. Note that $L(\theta; X)$ is largest when θ is as large as possible. However, the largest value of θ is equal to the smallest value of X_i in the sample. Thus, $\hat{\theta} = \min\{X_i\}\ 1 \leq i \leq n$.

Example 3.11
Let $X_1, X_2, \ldots, X_n$, denote a random sample from the normal distribution $N(\mu,\sigma^2)$. Then the likelihood function is given by.

$$L(X, \mu, \sigma^2) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

and

$$\ln L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Thus we have

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

Solving the two equations simultaneously, we obtain

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3.24}$$

Note that the MLEs, if they exist, are both sufficient and efficient estimates. They also have an additional property called invariance, i.e., for an MLE of θ, then $\mu\ (\theta)$ is the MLE of $\mu\ (\theta)$. However, they are not necessarily unbiased, i.e., $E(\hat{\theta}) = \theta$. The point in fact is σ^2:

$$E(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

Therefore, for small n, σ^2 is usually adjusted for its bias and the best estimate of σ^2 is

$$\hat{\sigma^2} = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Sometimes it is difficult, if not impossible, to obtain maximum likelihood estimators in a closed form, and therefore numerical methods must be used to maximize the likelihood function. For illustration see the following example.

Example 3.12
Suppose that X_1,X_2, …, X_n is a random sample from the Weibull distribution with pdf.

$$f(x, \alpha, \lambda) = \alpha\lambda x^{\alpha-1} e^{-\lambda x^{\alpha}} \tag{3.25}$$

The likelihood function is

$$L(X, \alpha, \lambda) = \alpha^n \lambda^n \left(\prod_{i=1}^{n} x_i^{\alpha-1}\right) e^{-\lambda \sum_{i=1}^{n} x_i^{\alpha}}$$

Then

$$\ln L = n \log \alpha + n \log \lambda + (\alpha - 1)\sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^{\alpha}$$

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log x_i - \lambda \sum_{i=1}^{n} x_i^{\alpha} \log x_i = 0$$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i^{\alpha} = 0$$

As noted, solutions of the above two equations for α and λ are extremely difficult and require either graphical or numerical methods.

Example 3.13
Let X_1, X_2, …, X_n be a random sample from the gamma distribution with pdf.

$$f(x, \lambda, \alpha) = \frac{\lambda^{(\alpha+1)} x^{\alpha} e^{-\lambda x}}{(\alpha!)^n} \tag{3.26}$$

then the likelihood function and log of the likelihood function, respectively, are

$$L(X, \lambda, \alpha) = \frac{\lambda^{n(\alpha+1)} \prod_{i=1}^{n} x_i^{\alpha}\ e^{-\lambda \sum_{i=1}^{n} x_i}}{(\alpha!)^n}$$

$$\ln L = n(\alpha+1)\log\lambda + \alpha\sum_{i=1}^{n}\log x_i - \lambda\sum_{i=1}^{n}x_i - n\log(\alpha!).$$

Taking the partial derivatives, we obtain

$$\begin{aligned}\frac{\partial \ln L}{\partial \alpha} &= n\log\lambda + \sum_{i=1}^{n}\log x_i - n\frac{\partial}{\partial\alpha}[\log\alpha!] = 0\\ \frac{\partial \ln L}{\partial \lambda} &= \frac{n(\alpha+1)}{\lambda} - \sum_{i=1}^{n}x_i = 0\end{aligned} \tag{3.27}$$

The solutions of the two equations at Eq. (3.27) for α and λ are extremely difficult and require either graphical or numerical methods.

Example 3.14
Let $t_1, t_2, \ldots, t_n$ be failure times of a random variable having the loglog distribution, also known as Pham distribution (Pham 2002), with two parameters a and α as follows (see also Eq. (2.82), Chap. 2):

$$f(t) = \alpha\ \ln(a)\ t^{\alpha-1}\ a^{t^{\alpha}}\ e^{1-a^{t^{\alpha}}} \quad \text{for } t>0, \alpha>0, a>1 \tag{3.28}$$

From Chap. 2, Eq. (2.83), the Pham cdf is given by:

$$F(t) = 1 - e^{1-a^{t^{\alpha}}}$$

We now estimate the values of a and α using the MLE method. From Eq. (3.28), the likelihood function is

$$\begin{aligned}L(a,\alpha) &= \prod_{i=1}^{n}\alpha\ln a\cdot t_i^{\alpha-1}e^{1-a^{t_i^{\alpha}}}a^{t_i^{\alpha}}\\ &= \alpha^n(\ln a)^n\left(\prod_{i=1}^{n}t_i\right)^{\alpha-1}a^{\sum_{i=1}^{n}t_i^{\alpha}}e^{n-\sum_{i=1}^{n}a^{t_i^{\alpha}}}\end{aligned}$$

The log likelihood function is

$$\begin{aligned}\log L(a,\alpha) &= n\log\alpha + n\ln(\ln a) + (\alpha-1)\left(\sum_{i=1}^{n}\ln t_i\right)\\ &\quad + \ln a\cdot\sum_{i=1}^{n}t_i^{\alpha} + n - \sum_{i=1}^{n}a^{t_i^{\alpha}}\end{aligned} \tag{3.29}$$

The first derivatives of the log likelihood function with respect to a and α are, respectively,

$$\frac{\partial}{\partial a}\log L(a,\alpha) = \frac{n}{a\ln a} + \frac{1}{a}\cdot\sum_{i=1}^{n} t_i^{\alpha} - \sum_{i=1}^{n} t_i^{\alpha} a^{t_i^{\alpha}-1} \tag{3.30}$$

and

$$\frac{\partial}{\partial \alpha}\log L(a,\alpha) = \frac{n}{\alpha} + \sum_{i=1}^{n}\ln t_i + \ln a\sum_{i=1}^{n}\ln t_i\ t_i^{\alpha} - \sum_{i=1}^{n} t_i^{\alpha} a^{t_i^{\alpha}}\ln a\ \ln t_i \tag{3.31}$$

Setting eqs. (3.30) and (3.31) equal to zero, we can obtain the MLE of a and α by solving the following simultaneous equations:

$$\frac{n}{\ln a} + \sum_{i=1}^{n} t_i^{\alpha} - \sum_{i=1}^{n} t_i^{\alpha} a^{t_i^{\alpha}} = 0$$
$$\frac{n}{\alpha} + \sum_{i=1}^{n}\ln t_i + \ln a\cdot\sum_{i=1}^{n}\ln t_i\cdot t_i^{\alpha}\left(1 - a^{t_i^{\alpha}}\right) = 0$$

After rearrangements, we obtain

$$\ln a\sum_{i=1}^{n} t_i^{\alpha}\left(a^{t_i^{\alpha}} - 1\right) = n$$
$$\ln a\cdot\sum_{i=1}^{n}\ln t_i\cdot t_i^{\alpha}\cdot\left(a^{t_i^{\alpha}} - 1\right) - \frac{n}{\alpha} = \sum_{i=1}^{n}\ln t_i$$

Example 3.15
Let $t_1, t_2, \ldots, t_n$ be failure times of a random variable having the vtub-shaped failure rate function. The probability density function $f(t)$ of its vtub-shaped failure rate is given by (see Eq. (1) in Chap. 2):

$$f(t) = \left(at\ln(bt) + \frac{a}{b}\right)\mathrm{e}^{-\left\{at\left[\frac{t}{2}\ln(bt) - \frac{t}{4} + \frac{1}{b}\right]\right\}} \quad \text{for } t > 0, a > 0, b > 0$$

We now estimate the two unknown parameters a and b using the MLE method. From the pdf above, the likelihood function is given by

$$L(a,b) = \prod_{i=1}^{n}\left(at_i\ln(bt_i) + \frac{a}{b}\right)\mathrm{e}^{-a\sum_{i=1}^{n} t_i\left[\frac{t_i}{2}\ln(bt_i) - \frac{t_i}{4} + \frac{1}{b}\right]}.$$

The log likelihood function is

$$\ln L(a,b) = \sum_{i=1}^{n} \ln\left(at_i \ln(bt_i) + \frac{a}{b}\right) - a\sum_{i=1}^{n} t_i\left[\frac{t_i}{2}\ln(bt_i) - \frac{t_i}{4} + \frac{1}{b}\right].$$

By taking the first derivatives of the log likelihood function above with respect to a and b and equate them to zero, and after some algebra, the MLE of a and b can be easily obtained by solving the following equations:

$$\begin{cases} a = \frac{2b\sum_{i=1}^{n} \frac{bx_i - 1}{(bx_i \ln(bx_i)+1)}}{\sum_{i=1}^{n} x_i(bx_i - 2)} \\ \sum_{i=1}^{n} \frac{bx_i \ln(bx_i)+1}{abx_i \ln(bx_i)+a} = \sum_{i=1}^{n} x_i\left[\frac{x_i}{2}\ln(bx_i) - \frac{x_i}{4} + \frac{1}{b}\right]. \end{cases}$$

Example 3.16
Suppose that $X_1, X_2, \ldots, X_n$ are independent random variable, each with the uniform distribution on $[a-d, a+d]$ where a is known and d is positive and unknown. Find the maximum likelihood estimator of d.

Solution: For $i = 1,2,\ldots, n$, the pdf of X_i is given by

$$f(d, x_i) = \begin{cases} \frac{1}{2d} & \text{if a} - \text{d} \le \text{x}_\text{i} \le a + d \\ 0 & \text{otherwise.} \end{cases} \tag{3.32}$$

The likelihood function is

$$L(d) = \begin{cases} \left(\frac{1}{2d}\right)^n & \text{if a} - \text{d} \le x_1, x_2, \ldots, x_n \le a + d \\ 0 & \text{otherwise.} \end{cases}$$

To maximize L(c,d) is the same as to minimize $(2d)^n$. In other words, L(d) will be maximized by the smallest possible d with L(d) > 0. This implies for all $i = 1,2,\ldots,n$ $d \ge -x_i + \text{a}$ and $d \ge x_i - \text{a}$. Thus, $\hat{d} = \max_{1 \le i \le n} \{|x_i - a|\}$.

3.4 Invariance and Lower Bound on the Variance

In this section we discuss some properties of MLEs and how to establish a lower bound on the variance using an inequality known as the Cramér-Rao inequality.

Theorem 3.1 (Invariance Principle)
If $\hat{\theta}$ is the MLE of parameter θ then $g(\hat{\theta})$ is the MLE of parameter $g(\theta)$.

In other words, if the likelihood L(X, θ) *is maximized at the point.* $\hat{\theta} = h(x_1, x_2, \ldots, x_n)$ *then the function* L(X, $g(\theta)$) *is maximized at the point*

$$g\left(\hat{\theta}\right) = g(h(x_1, x_2, \ldots, x_n)).$$

Example 3.17
If $\hat{\theta}$ is the MLE of the variance σ^2 then $\sqrt{\hat{\theta}}$ is the MLE of the standard deviation σ.

Theorem 3.2 (Likelihood Principle) *Consider two sets of data,* x *and* y, *obtained from the same population, although possibly according to different sampling plans. If the ratio of their likelihoods,* $\frac{L_1(x,\theta)}{L_2(y,\theta)}$, *does not depend on* θ, *then both data sets provide the same information about the parameter* θ *and consequently should lead to the same conclusion about* θ.

Example 3.18
Assume that we want to estimate θ, the probability of success of Bernoulli trials. One experimental design consists of fixing n, the number of observations, and recording the number of successes. If we observe x successes, the likelihood is.

$$L_1(x, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

Suppose that one decide to fix x and take observations until x successes are recorded. The probability that the observations will end on the nth trial is given by negative binomial distribution, and the likelihood is

$$L_2(n, \theta) = \binom{n-1}{x-1} \theta^x (1-\theta)^{n-x}.$$

Note that in the first case x was random and n fixed; in the second it is the other way around. Thus, the ratio of these two likelihoods is given by

$$\frac{L_1(x, \theta)}{L_2(n, \theta)} = \frac{\binom{n}{x}}{\binom{n-1}{x-1}},$$

which does not depend on the parameter θ. The MLE of θ is $\hat{\theta} = \frac{x}{n}$ in either case. Hence the additional information that in the second case the last experiment led to success does not affect our estimate of the parameter θ.

Theorem 3.3 (Cramér-Rao inequality) *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with pdf $f(x; \theta)$ for $\theta_1 < \theta < \theta_2$, where θ_1 and θ_2 are known. Let $Y = h(X_1, X_2, \ldots, X_n)$ be an unbiased estimator of θ. The lower bound inequality on the variance of Y, Var(Y), is given by.*

$$Var(Y) \geq \frac{1}{nE\left\{\left[\frac{\partial \ln f(x;\theta)}{\partial \theta}\right]^2\right\}} = -\frac{1}{nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)} \tag{3.33}$$

or

$$Var(Y) \geq \frac{1}{n\, I(\theta)} \tag{3.34}$$

where

$$I(\theta) \;=\; E\left\{\left[\frac{\partial \ln f(x;\theta)}{\partial \theta}\right]^2\right\}$$

also known as Fisher's Information (see Problem 15).

Theorem 3.4 *An estimator $\hat{\theta}$ is said to be asymptotically efficient if $\sqrt{n}\hat{\theta}$ has a variance that approaches the Cramér-Rao lower bound for large n, that is,*

$$\lim_{n\to\infty} Var(\sqrt{n}\hat{\theta}) = -\frac{1}{nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)}. \tag{3.35}$$

Example 3.19
Let $X_1, X_2, \ldots, X_n$ denote a random sample from the Bernoulli distribution with a pdf.

$$f(x) = p^x(1-p)^{1-x} \qquad x = 0, 1 \qquad 0 \leq p \leq 1$$

where p is the probability of success in each trial.

Let Y represents the number of successes in n independent trials, that is, $Y = \sum_{i=1}^{n} X_i$ then we can easily show that $W = \frac{Y}{n}$ is an unbiased estimator of p. We now determine if W is the minimum variance unbiased estimator of p. The variance of W is given by

$$V(W) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

To obtain the lower bound of the Cramer-Rao inequality, we need to compute the following

$$\ln f = x \ln p + (1-x)\ln(1-p)$$

$$\frac{\partial}{\partial p}\ln f = \frac{x}{p} - \frac{(1-x)}{(1-p)}$$
$$\frac{\partial^2}{\partial p^2}\ln f = -\frac{x}{p^2} - \frac{(1-x)}{(1-p)^2}.$$

Then

$$E\left\{\frac{\partial^2}{\partial p^2}\ln f\right\} = -\frac{p}{p^2} - \frac{(1-p)}{(1-p)^2} = -\frac{1}{p(1-p)}.$$

Thus,

$$Var(W) \geq -\frac{1}{nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)}$$
$$= -\frac{1}{n\left(-\frac{1}{p(1-p)}\right)} = \frac{p(1-p)}{n} = Var(W).$$

This means that W is the minimum variance unbiased estimator of p.

Example 3.20
Let T_1,T_2, …, T_n denote a random sample from an exponential distribution with the mean μ and its pdf is given by.

$$f(t) = \frac{1}{\mu}e^{-\frac{t}{\mu}} \qquad t \geq 0,\ \mu > 0.$$

Note that, from example 3.5, the MLE of μ is $\widehat{\mu} = \frac{\sum_{i=1}^{n} T_i}{n}$. Define $Y = \frac{\sum_{i=1}^{n} T_i}{n}$. Is Y the minimum variance unbiased estimator of μ?

The variance of Y is given by

$$Var(Y) = Var\left(\frac{\sum_{i=1}^{n} T_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(T_i) = \frac{1}{n^2}\left(n\mu^2\right) = \frac{\mu^2}{n}.$$

We now compute the lower bound using the Cramer-Rao inequality.

$$\ln f = -\ln \mu - \frac{t}{\mu}$$
$$\frac{\partial \ln f}{\partial \mu} = -\frac{1}{\mu} + \frac{t}{\mu^2}$$

$$\frac{\partial^2 \ln f}{\partial \mu^2} = \frac{1}{\mu^2} - \frac{2t}{\mu^3}.$$

Then

$$E\left\{\frac{\partial^2 \ln f}{\partial \mu^2}\right\} = \frac{1}{\mu^2} - \frac{2E(T)}{\mu^3} = \frac{1}{\mu^2} - \frac{2\mu}{\mu^3} = -\frac{1}{\mu^2}.$$

Thus,

$$\begin{aligned} Var(Y) &\geq -\frac{1}{nE\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)} \\ &= -\frac{1}{n\left(-\frac{1}{\mu^2}\right)} = \frac{\mu^2}{n} = Var(Y). \end{aligned}$$

This shows that Y is the minimum variance unbiased estimator of μ.

3.5 Maximum Likelihood Estimation with Censored Data

Censored data arises when an individual's life length is known to occur only in a certain period of time. In other words, a censored observation contains only partial information about the random variable of interest. In this section, we consider two types of censoring. The first type is called Type-I censoring where the event is observed only if it occurs prior to some pre-specified time. The second type is Type-II censoring in which the study continues until the failure of the first r units (or components), where r is some predetermined integer ($r < n$).

Examples of Type-II censoring are often used in testing of equipment life. Here items are put on test at the same time, and the test is terminated when r of the n items have failed and without replacement. Such an experiment may, however, save time and resources because it could take a very long time for all items to fail. Both Type-I and Type-II censoring arise in many reliability applications.

For example, there is a batch of transistors or tubes; we put them all on test at $t = 0$, and record their times to failure. Some transistors may take a long time to burn out, and we will not want to wait that long to end the experiment. Therefore, we might stop the experiment at a pre-specified time t_c, in which case we have Type-I censoring, or we might not know beforehand what value of the fixed censoring time is good so we decide to wait until a pre-specified number of units have failed, r, of all the transistors has burned out, in which case we have Type-II censoring.

Censoring times may vary from individual to individual or from application to application. We now discuss a generalized censoring times case, call a multiple-censored data.

3.5.1 Parameter Estimate with Multiple-Censored Data

A sample of n units are drawn at random and put on test with a pdf f and cdf F. The likelihood function for the multiple-censored data is given by

$$L = f(t_{1,f}, \ldots, t_{r,f}, t_{1,s}, \ldots, t_{m,s}) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(t_{i,f}) \prod_{j=1}^{m} [1 - F(t_{j,s})] \quad (3.36)$$

where $f(.)$ is the density function and $F(.)$ is the distribution function. There are r failures at times $t_{1,f}, \ldots, t_{r,f}$ and m units (here, $m = n$ s$- r$) with censoring times $t_{1,s}, \ldots, t_{m,s}$. Note that this includes Type-I censoring by simply setting $t_{\mathrm{i,f}} = \mathrm{t_{i,\,n}}$ and $t_{\mathrm{j,s}} = t_0$ in the likelihood function in Eq. (3.36). Also, the likelihood function for Type-II censoring is similar to Type-I censoring except $t_{\mathrm{j,s}} = t_{\mathrm{r}}$ in Eq. (3.36). In other words, the likelihood function for the first r observations from a sample size n drawn from the model in both Type-I and Type-II censoring is given by

$$L = f(t_{1,n}, \ldots, t_{r,n}) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(t_{i,n}) \, [1 - F(t_*)]^{n-r} \quad (3.37)$$

where $t_* = t_0$, the time of cessation of the test for Type-I censoring and $t_* = t_r$, the time of the rth failure for Type-II censoring.

Example 3.21
Consider a two-parameter probability density distribution with multiple-censored data and distribution function with failure rate bathtub shape, as given by (Chen 2000):

$$f(t) = \lambda\beta t^{\beta-1} \exp\left[t^{\beta} + \lambda(1 - e^{t^{\beta}})\right], \qquad t, \lambda, \beta > 0 \quad (3.38)$$

and

$$F(t) = 1 - \exp\left[\lambda(1 - e^{t^{\beta}})\right], \qquad t, \lambda, \beta > 0 \quad (3.39)$$

respectively. Substituting the functions $f(t)$ and $F(t)$ in Eqs. (3.38) and (3.39) into Eq. (3.37), we obtain the logarithm of the likelihood function:

$$\ln L = \ln \frac{n!}{(n-r)!} + r \ln \lambda + r \ln \beta + \sum_{i=1}^{r} (\beta - 1) \ln t_i + (m + r)\lambda + \sum_{i=1}^{r} t_i^{\beta} - \left(\sum_{i=1}^{r} \lambda e^{t_i^{\beta}} + \sum_{j=1}^{m} \lambda e^{t_j^{\beta}} \right). \quad (3.40)$$

The function $\ln L$ can be maximized by setting the partial derivative of $\ln L$ with respect to λ and β, equal to zero and solving the resulting equations simul-taneously for λ and β. Therefore, we obtain

$$\frac{\partial \ln L}{\partial \lambda} = \frac{r}{\lambda} + (m+r) - \sum_{i=1}^{r} e^{t_i^{\beta}} - \sum_{j=1}^{m} e^{t_j^{\beta}} \equiv 0$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{r}{\beta} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\beta} \ln t_i - \lambda \left[\sum_{i=1}^{r} e^{t_i^{\beta}} t_i^{\beta} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\beta}} t_j^{\beta} \ln t_j \right] \equiv 0.$$

This implies that

$$\hat{\lambda} = \frac{r}{\left(\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} \right) - m - r} \tag{3.41}$$

and $\hat{\beta}$ is the solution of

$$\frac{r}{\hat{\beta}} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\hat{\beta}} \ln t_i = \frac{r}{\left(\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} \right) - m - r} \left[\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} t_j^{\hat{\beta}} \ln t_j \right] \tag{3.42}$$

We now discuss two special cases as follows.

Case I: Type-I or Type-II Censoring Data.

From Eq. (3.37), the likelihood function for the first r observations from a sample size n drawn from the model in both Type-I and Type-II censoring is

$$L = f(t_{1,n}, \ldots, t_{r,n}) = \frac{n!}{(n-r)!} \prod_{i=1}^{r} f(t_{i,n}) \left[1 - F(t_*)\right]^{n-r}$$

where $t_* = t_0$, the time of cessation of the test for Type-I censoring and $t_* = t_r$, the time of the rth failure for Type-II censoring Eqs. (3.41) and (3.42) become

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + (n-r) e^{t_*^{\hat{\beta}}} - n}. \tag{3.43}$$

$$\frac{r}{\hat{\beta}} + \sum_{i=1}^{r} \ln t_i + \sum_{i=1}^{r} t_i^{\hat{\beta}} \ln t_i$$

$$= \frac{r}{\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} + (n-r)e^{t_*^{\hat{\beta}}} - n} \left[\sum_{i=1}^{r} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i + \sum_{j=1}^{m} e^{t_j^{\hat{\beta}}} t_j^{\hat{\beta}} \ln t_j \right]. \tag{3.44}$$

Case II: Complete Censored Data.

Simply replace r with n in Eqs. (3.41) and (3.42) and ignore the t_j portions. The maximum likelihood equations for the λ and β are given by

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} - n} \tag{3.45}$$

$$\frac{n}{\hat{\beta}} + \sum_{i=1}^{n} \ln t_i + \sum_{i=1}^{n} t_i^{\hat{\beta}} \ln t_i = \frac{n}{\sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} - n} \times \sum_{i=1}^{n} e^{t_i^{\hat{\beta}}} t_i^{\hat{\beta}} \ln t_i. \tag{3.46}$$

3.5.2 Confidence Intervals of Estimates

The asymptotic variance-covariance matrix of the parameters (λ and β) is obtain- ed by inverting the Fisher information matrix

$$I_{ij} = E\left[-\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, 2 \tag{3.47}$$

where $\theta_1, \theta_2 = \lambda$ or β (Nelson et al.1992). This leads to

$$\begin{bmatrix} Var(\hat{\lambda}) & Cov(\hat{\lambda}, \hat{\beta}) \\ Cov(\hat{\lambda}, \hat{\beta}) & Var(\hat{\beta}) \end{bmatrix}$$

$$= \begin{bmatrix} E\left(-\frac{\partial^2 \ln L}{\partial^2 \lambda} |_{\hat{\lambda}, \hat{\beta}} \right) & E\left(-\frac{\partial^2 \ln L}{\partial \lambda \partial \beta} |_{\hat{\lambda}, \hat{\beta}} \right) \\ E\left(-\frac{\partial^2 \ln L}{\partial \beta \partial \lambda} |_{\hat{\lambda}, \hat{\beta}} \right) & E\left(-\frac{\partial^2 \ln L}{\partial^2 \beta} |_{\hat{\lambda}, \hat{\beta}} \right) \end{bmatrix}^{-1} \tag{3.48}$$

We can obtain an approximate (1–α)100% confidence intervals on parameter λ and β based on the asymptotic normality of the MLEs (Nelson et al. 1992) as follows:

$$\hat{\lambda} \pm Z_{\alpha/2} \sqrt{Var(\hat{\lambda})} \quad \text{and} \quad \hat{\beta} \pm Z_{\alpha/2} \sqrt{Var(\hat{\beta})} \tag{3.49}$$

where $Z_{\alpha/2}$ is upper percentile of standard normal distribution.

Equation Sect. 1.

3.5.3 Applications

Consider a helicopter main rotor blade part code xxx-015-001-107 based on the system database collected from October 1995 to September 1999 (Pham 2002). The data set is shown in Table 3.1. In this application, we consider several distribution functions including Weibull, lognormal, normal, and loglog distribution functions. From Example 3.14, the Pham pdf (see Eq. 3.28) with parameters a and α is.

$f(t) = \alpha \cdot \ln a \cdot t^{\alpha-1} \cdot a^{t^\alpha} \cdot e^{1-a^{t^\alpha}}$ for $t > 0, \alpha > 0, a > 0$

and its corresponding log likelihood function (see Eq. 3.30) is

$$\log L(a,\alpha) = n\log\alpha + n\ln(\ln a) + (\alpha-1)\left(\sum_{i=1}^{n}\ln t_i\right) + \ln a \cdot \sum_{i=1}^{n} t_i^\alpha + n - \sum_{i=1}^{n} a^{t_i^\alpha}$$

We next determine the confidence intervals for parameter estimates a and α. For the log-likelihood function given in Eq. (3.30), we can obtain the Fisher information matrix H as $H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$ where $h_{11} = E\left[-\frac{\partial^2 \log L}{\partial a^2}\right]$

Table 3.1 Main rotor blade data (hour)

1634.3	2094.3	3318.2
1100.5	2166.2	2317.3
1100.5	2956.2	1081.3
819.9	795.5	1953.5
1398.3	795.5	2418.5
1181	204.5	1485.1
128.7	204.5	2663.7
1193.6	1723.2	1778.3
254.1	403.2	1778.3
3078.5	2898.5	2943.6
3078.5	2869.1	2260
3078.5	26.5	2299.2
26.5	26.5	1655
26.5	3180.6	1683.1
3265.9	644.1	1683.1
254.1	1898.5	2751.4
2888.3	3318.2	
2080.2	1940.1	

$$h_{12} = h_{21} = E\left[-\frac{\partial^2 \log L}{\partial a\, \partial \alpha}\right] \quad \text{and} \quad h_{22} = E\left[-\frac{\partial^2 \log L}{\partial \alpha^2}\right]$$

The variance matrix, V, can be obtained as follows:

$$V = [H]^{-1} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \tag{3.50}$$

The variances of a and α are

$$Var(a) = v_{11} \quad Var(\alpha) = v_{22}$$

One can approximately obtain the 100(1–β)% confidence intervals for a and α based on the normal distribution as $[\hat{a} - z_{\frac{\beta}{2}}\sqrt{v_{11}}, \hat{a} + z_{\frac{\beta}{2}}\sqrt{v_{11}}]$ and $[\hat{\alpha} - z_{\frac{\beta}{2}}\sqrt{v_{22}}, \hat{\alpha} + z_{\frac{\beta}{2}}\sqrt{v_{22}}]$, respectively, where v_{ij} is given in Eq. (3.50) and z_β is (1-β/2)100% of the standard normal distribution. After we obtain $\hat{a}$ and $\hat{\alpha}$, the MLE of reliability function can be computed as

$$\hat{R}(t) = e^{1-\hat{a}^{t^{\hat{\alpha}}}}$$

Let us define a partial derivative vector for reliability $R(t)$ as

$$v[R(t)] = \left[\frac{\partial R(t)}{\partial a} \quad \frac{\partial R(t)}{\partial \alpha}\right]$$

then the variance of $R(t)$ can be obtained as follows:

$$Var[R(t)] = v[R(t)] \cdot V \cdot (v[R(t)])^T \tag{3.51}$$

where V is given in Eq. (3.50).

One can approximately obtain the (1−β)100% confidence interval for $R(t)$ as

$$\left[\hat{R}(t) - z\beta\sqrt{Var[R(t)]}, \hat{R}(t) + z\beta\sqrt{Var[R(t)]},\right].$$

The MLE parameter estimations of Pham distribution using the data set in Table 3.1 are given as follows:

$\hat{\alpha} = 1.1075 \quad \text{Var}[\hat{\alpha}] = 0.0162$
95% CI for $\hat{\alpha}$: [0.8577, 1.3573]
$\hat{a} = 1.0002 \quad \text{Var}[\hat{a}] = 2.782e^{-8}$
95%CI for a: [0.9998, 1.0005]
MTTF = 1608.324 MRL(t = MTTF) = 950.475

Figure 3.1 shows the loglog reliability and its 95% confidence interval of main rotor blade, respectively. Figure 3.2 shows the reliability comparisons between the normal, the lognormal, Weibull, and the loglog distributions for the main rotor blade data set.

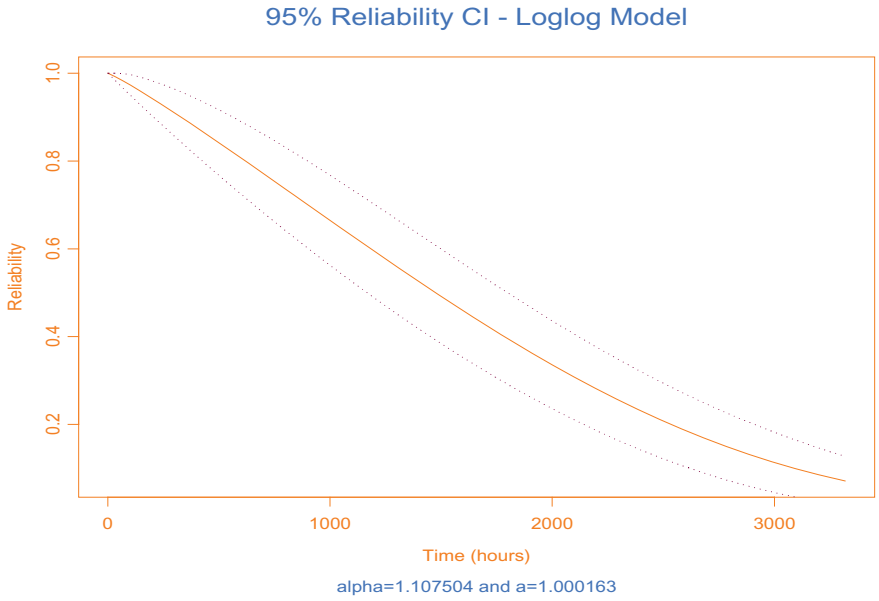

Fig. 3.1 Estimated reliability and its confidence interval for a main rotor blade data set

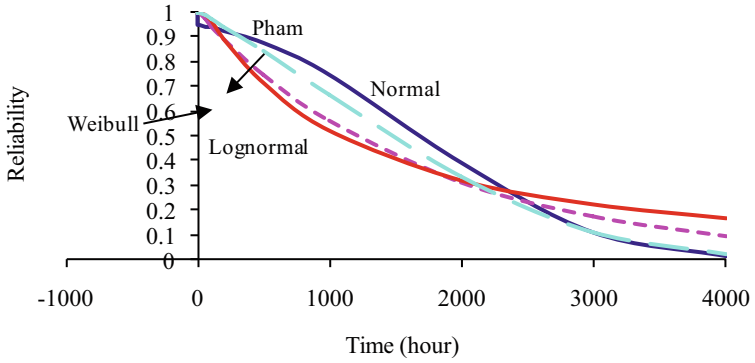

Fig. 3.2 Reliability comparisons for a main rotor blade data set

3.6 Statistical Change-Point Estimation Methods

The change-point problem has been widely studied in reliability applications such as biological sciences, survival analysis, and environmental statistics.

Assume there is a sequence of random variables $X_1, X_2, \ldots, X_n$, that represent the inter-failure times and exists an index change-point τ, such that $X_1, X_2, \ldots, X_\tau$ have a common distribution F with density function $f(t)$ and $X_{\tau+1}, X_{\tau+2}, \ldots, X_n$ have the distribution G with density function $g(t)$, where $F \neq G$. Consider the following assumptions:

1. There is a finite unknown number of units, N, to put under the test.
2. At the beginning, all of the units have the same lifetime distribution F. After τ failures are observed, the remaining $(N-\tau)$ items have the distribution G. The change-point τ is assumed unknown.
3. The sequence $\{X_1, X_2, \ldots, X_\tau\}$ is statistically independent of the sequence $\{X_{\tau+1}, X_{\tau+2}, \ldots, X_n\}$.
4. The lifetime test is performed according to the Type-II censoring plan in which the number of failures, n, is pre-determined.

Note that in hardware reliability testing, the total number of units to put on the test N can be determined in advance. But in software, the parameter N can be defined as the initial number of faults, and therefore it makes more sense for it to be an unknown parameter. Let $T_1, T_2, \ldots, T_n$ be the arrival times of sequential failures. Then

$$\begin{aligned} T_1 &= X_1 \\ T_2 &= X_1 + X_2 \\ &\vdots \\ T_n &= X_1 + X_2 + \ldots X_n \end{aligned} \tag{3.52}$$

The failure times $T_1, T_2, \ldots, T_\tau$ are the first τ order statistics of a sample of size N from the distribution F. The failure times $T_{\tau+1}, T_{\tau+2}, \ldots, T_n$ are the first $(n-\tau)$ order statistics of a sample of size $(N-\tau)$ from the distribution G.

Example 3.22
The Weibull change-point model of given life time distributions F and G with parameters (λ_1, β_1) and (λ_2, β_2), respectively, can be expressed as follows:

$$F(t) = 1 - \exp\left(-\lambda_1 t^{\beta_1}\right) \tag{3.53}$$

$$G(t) = 1 - \exp\left(-\lambda_2 t^{\beta_2}\right). \tag{3.54}$$

Assume that the distributions belong to parametric families $\{F(t\,|\theta_1)\,, \theta_1 \in \Theta_1\}$ and $\{G(t\,|\theta_2)\,, \theta_2 \in \Theta_2\}$. Assume $T_1, T_2, \ldots, T_\tau$ are the first τ order statistics of a sample with size N from the distribution $\{F(t\,|\theta_1)\,, \theta_1 \in \Theta_1\}$ and $T_{\tau+1}, T_{\tau+2}, \ldots, T_n$

are the first $(n-\tau)$ order statistics of a sample of size $(N-\tau)$ from the distribution $\{G(t\,|\,\theta_2\,),\ \theta_2 \in \Theta_2\}$ where N is unknown. The log likelihood function can be expressed as follows (Zhao 2003):

$$\begin{aligned} &L(\tau, N, \theta_1, \theta_2 | T_1, T_2, \ldots, T_n) \\ &= \sum_{i=1}^{n} (N - i + 1) + \sum_{i=1}^{\tau} f(T_i|\theta_1) \\ &+ \sum_{i=\tau+1}^{n} g(T_i|\theta_2) + (N - \tau) \log(1 - F(T_\tau|\theta_1)) \\ &+ (N - n) \log(1 - G(T_n|\theta_2)). \end{aligned} \tag{3.55}$$

If the parameter N is known where hardware reliability is commonly considered, then the likelihood function is given by

$$\begin{aligned} &L(\tau, \theta_1, \theta_2 | T_1, T_2, \ldots, T_n) \\ &= \sum_{i=1}^{\tau} f(T_i|\theta_1) + \sum_{i=\tau+1}^{n} g(T_i|\theta_2) \\ &+ (N - \tau) \log(1 - F(T_\tau|\theta_1)) + (N - n) \log(1 - G(T_n|\theta_2)). \end{aligned}$$

The MLE of the change-point value $\hat{\tau}$ and $(\hat{N}, \hat{\theta}_1, \hat{\theta}_2)$ can be obtained by taking partial derivatives of the log likelihood function in Eq. (3.55) with respect to the unknown parameters that maximizes the function. It should be noted that there is no closed form for $\hat{\tau}$ but it can be obtained by calculating the log likelihood for each possible value of τ, $1 \leq \tau \leq (n-1)$, and selecting as $\hat{\tau}$ the value that maximizes the log-likelihood function.

3.6.1 *Application: A Software Model with a Change Point*

In this application we examine the case where the sample size N is unknown. Consider a software reliability model developed by Jelinski and Moranda (1972), often called the Jelinski-Moranda model. The assumptions of the model are as follows:

1. There are N initial faults in the program.
2. A detected fault is removed instantaneously and no new fault is introduced.
3. Each failure caused by a fault occurs independently and randomly in time according to an exponential distribution.
4. The functions F and G are exponential distributions with failure rate parameters λ_1 and λ_2, respectively.

Based on the assumptions, the inter-failure times $X_1, X_2, \ldots, X_n$ are independently exponentially distributed. Specifically,$X_i = T_i - T_{i-1}, i = 1, 2, \ldots \tau$, are exponen-tially distributed with parameter $\lambda_1(N - i + 1)$ where λ_1 is the initial fault detection rate of the first τ failures and $X_j = T_j - T_{j-1}$, $j = \tau + 1, \tau + 2, \ldots n$, are exponentially distributed with parameter $\lambda_2(N - \tau - j + 1)$ where λ_2 is the fault detection rate of the first $(n - \tau)$ failures. If $\lambda_1 = \lambda_2$ it means that each fault removal is the same and the change-point model becomes the Jelinski-Moranda software reliability model (Jelinski and Moranda 1972).

The MLEs of the parameters $(\tau, N, \lambda_1, \lambda_2)$ can be obtained by solving the following equations simultaneously:

$$\hat{\lambda}_1 = \frac{\tau}{\sum_{i=1}^{\tau} \left(\hat{N} - i + 1\right) x_i} \tag{3.56}$$

$$\hat{\lambda}_2 = \frac{(n - \tau)}{\sum_{i=\tau+1}^{n} \left(\hat{N} - i + 1\right) x_i} \tag{3.57}$$

$$\sum_{i=1}^{n} \frac{1}{(\hat{N} - i + 1)} = \hat{\lambda}_1 \sum_{i=1}^{\tau} x_i + \hat{\lambda}_2 \sum_{i=\tau+1}^{n} x_i. \tag{3.58}$$

To illustrate the model, we use the data set as in Table 3.2 to obtain the unknown parameters $(\tau, N, \lambda_1, \lambda_2)$ using Eqs. (3.56)–(3.58). The data in Table 3.2 (Musa et al. 1987) shows the successive inter-failure times for a real-time command and control system. The table reads from left to right in rows, and the recorded times are execution

Table 3.2 Successive inter-failure times (in seconds) for a real-time command system

3	30	113	81	115	9	2	91	112	15
138	50	77	24	108	88	670	120	26	114
325	55	242	68	422	180	10	1146	600	15
36	4	0	8	227	65	176	58	457	300
97	263	452	255	197	193	6	79	816	1351
148	21	233	134	357	193	236	31	369	748
0	232	330	365	1222	543	10	16	529	379
44	129	810	290	300	529	281	160	828	1011
445	296	1755	1064	1783	860	983	707	33	868
724	2323	2930	1461	843	12	261	1800	865	1435
30	143	108	0	3110	1247	943	700	875	245
729	1897	447	386	446	122	990	948	1082	22
75	482	5509	100	10	1071	371	790	6150	3321
1045	648	5485	1160	1864	4116				

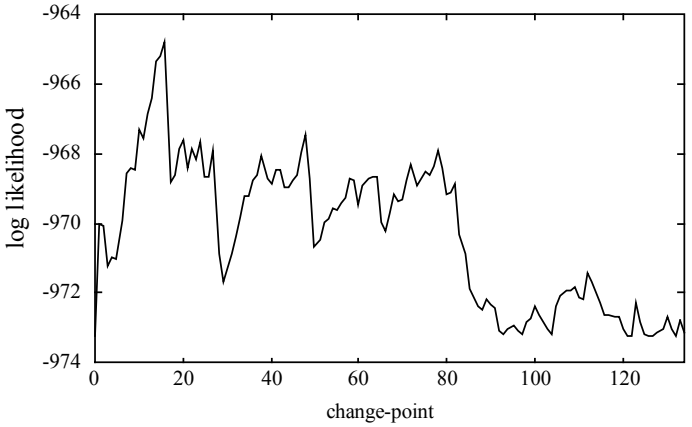

Fig. 3.3 The log-likelihood function versus the number of failures

times, in seconds. There are 136 failures in total. Figure 3.3 plots the log-likelihood function *vs* number of failures. The MLEs of the parameters $(\tau, N, \lambda_1, \lambda_2)$ with one change-point are given by

$$\hat{\tau} = 16, \hat{N} = 145, \hat{\lambda}_1 = 1.1 \times 10^{-4}, \hat{\lambda}_2 = 0.31 \times 10^{-4}.$$

If we do not consider a change-point in the model, the MLEs of the parameters N and λ can be given as

$$\hat{N} = 142, \hat{\lambda} = 0.35 \times 10^{-4}.$$

From Fig. 3.3, we can observe that it is worth considering the change-points in the reliability functions.

3.7 Goodness of Fit Tests

The problem at hand is to compare some observed sample distribution with a theoretical distribution. In fact, a practitioner often wonders how to test some hypothesis about the distribution of a population. If the test is concerned with the agreement between the distribution of a set of observed sample values and a theoretical distribution, we call it a “test of goodness of fit".

The basic question in validating distribution models is whether the shape of the fitted model corresponds to that of the data. To do that, we may just simply make a direct comparison of the observed data with what we expect to see from the fitted distribution model. In this section, two common tests of goodness of fit that will be discussed are the Chi-square test, χ^2, and the Kolmogorov-Smirnov (KS) test.

Note that the chi-square test requires all cell frequencies to exceed 5, whereas this restriction is no longer required for KS test because there is no need to classify the observations in carrying out the KS test.

3.7.1 The Chi-Square Test

The chi-square test often requires large samples and is applied by comparing the observed frequency distribution of the sample to the expected value under the assumption of the distribution. More specifically, consider a large sample of size N. Let $a_0 < a_1 < \ldots < a_k$ be the upper points of k subintervals of the frequency distribution. Basically, the statistic

$$\chi^2 = \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \tag{3.59}$$

has a chi-square (χ^2) distribution with k degrees of freedom where μ_i and σ_i.

are the mean and the standard deviation from the normal distribution in the ith subinterval, $i = 1,2,\ldots, k$.

Let f_i and E_i, for $i = 1, 2,\ldots, k$, be the observed frequency and the expected frequency in the ith subinterval, respectively. The expected frequency E_i in the ith subinterval is.

$E_i = N (F_i - F_{i-1})$.

The chi-square test statistic is defined as

$$S = \sum_{i=1}^{k} \frac{(f_i - E_i)^2}{E_i}. \tag{3.60}$$

It should be noted that

$$\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} E_i = N$$

where N is the sample size. Therefore, the chi-square statistic S can be rewritten as

$$S = \sum_{i=1}^{k} \frac{f_i^2}{E_i} - N. \tag{3.61}$$

The value of S is approximately distributed of $\chi^2_{1-\alpha,k-1}$. Thus, if $S \geq \chi^2_{1-\alpha,k-1}$. then the distribution $F(x)$ does not fit the observed data.

If the values of the parameters (some unknown parameters) of the distribution have to be estimated from the sample, then we need to reduce the number of degrees of freedom of χ^2 by the number of estimated parameters. In other words, the degrees of freedom are:

(the number of frequencies – 1 – number of parameters estimated with the same data).

In this case, the distribution $F(x)$ does not fit the observed data if $S \geq \chi^2_{1-\alpha,k-1-r}$ where r is the number of estimated parameters.

It should be noted that in order to get a good approximation the sample size N needs to be large and values of E_i should not be small, not much less than 5. The steps of chi-squared goodness of fit test are as follows:

1. Divide the sample data into the mutually exclusive cells (normally 8–12) such that the range of the random variable is covered.
2. Determine the frequency, f_i, of sample observations in each cell.
3. Determine the theoretical frequency, E_i, for each cell (area under density function between cell boundaries X_n—total sample size). Note that the theoretical frequency for each cell should be greater than 1. To carry out this step, it normally requires estimates of the population parameters which can be obtained from the sample data.
4. Form the statistic

$$S = \sum_{i=1}^{k} \frac{(f_i - E_i)^2}{E_i}$$

where $E_\mathrm{i} = \mathrm{N}\,(\mathrm{F_i} - \mathrm{F_{i-1}})$.

5. From the χ^2 tables, choose a value of χ^2 with the desired significance level and with degrees of freedom (=k–r–1), where r is the number of population parameters estimated.
6. Reject the hypothesis that the sample distribution is the same as theoretical distribution if

$$S \geq \chi^2_{1-\alpha,\, k-r-1}$$

where α is called the significance level.

Example 3.23

Given the data in Table 3.3, can the data be represented by the exponential distribution with a significance level of α?

Solution: From the above calculation, $\hat{\lambda} = 0.00263$, $R_i = e^{-\lambda t_i}$ and $F_i = 1-R_i$. Given that a value of significance level α is 0.1, from Eq. (3.60) we obtain

Table 3.3 Sample observations in each cell boundary ($N = 60$)

Cell boundaries	f_i	$Q_i = Fi * 60$	$E_i = \mathrm{N}(\mathrm{F_i} - \mathrm{F_{i-1}}) = Q_i - Q_{i-1}$
0–100	10	13.8755	13.8755
100–200	9	24.5422	10.6667
200–300	8	32.7421	8.1999
300–400	8	39.0457	6.3036
400–500	7	43.8915	4.8458
500–600	6	47.6168	3.7253
600–700	4	50.4805	2.8637
700–800	4	52.6819	2.2010
800–900	2	54.3743	1.6924
900–1,000	1	55.6753	1.3010
>1,000	1	60.0000	4.3247

$$S = \sum_{i=1}^{11} \frac{(f_i - E_i)^2}{E_i} = 8.7332$$

Here $k = 11$, $r = 1$. Then k–r–1 = 11–1–1 = 9. From Table A.3 in Appendix A, the value of χ^2 with nine degrees of freedom is 14.68, that is,

$$\chi^2_{9df}(0.90) = 14.68$$

Since $S = 8.7332 < 14.68$, we would not reject the hypothesis of exponential with $\lambda = 0.00263$.

Example 3.24

The number of defective soldering points found on a circuit board in a given day is given in Table 3.4. Now the question here is: Can the data be represented by the Poisson distribution with a significance level of α?

Solution: From Table 3.4, one can estimate the rate: $\hat{\lambda} = 56/33 = 1.7$. The expected frequency $\mathrm{E_i}$ is given by:

Table 3.4 Sample observations in each cell boundary ($N = 33$)

Number of defective	Observed frequency f_i	Expected frequency E_i	Total number of defectives
0	6	6.04	0
1	8	10.26	8
2	12	8.71	24
3	4	4.92	12
4 or more	3	3.07	12
Total	33		56

$$E_i = NP(D = i) = \mathrm{N}\frac{e^{-\lambda}\lambda^{\mathrm{i}}}{\mathrm{i!}}.$$

Note that.
$P(D = d) = \frac{e^{-\lambda}\lambda^{\mathrm{d}}}{\mathrm{d!}}$ for d = 0, 1, …, 4
Given that a value of significance level α is 0.1, from Eq. (3.60) we obtain

$$S = \sum_{i=0}^{4} \frac{(f_i - E_i)^2}{E_i} = 1.912$$

From Table A.3 in Appendix A, the value of χ^2 with three degrees of freedom ($k-1-1 = 3$) is

$$\chi^2_{3df}(0.90) = 6.25$$

Since $S = 1.912 < 6.25$, we would not reject the hypothesis of Poisson model.

Example 3.25
Consider a sample of 100 failure times of electric generators (in hours). We now determine whether the distribution of failure times is normal. Table 3.5 gives the observed frequencies and expected frequencies over 8 intervals. Given that the estimated values of the mean μ and standard deviation σ are $\mu = 0.122$ and $\sigma = 0.011$.

we obtain

$$S = \sum_{i=1}^{8} \frac{(f_i - E_i)^2}{E_i} = 5.404$$

From Table A.3 in Appendix A, the value of χ^2 with 5 degrees of freedom (i.e., $k-1-r = 8-1-2 = 5$) is

Table 3.5 Sample observations of 100 failure times ($N = 100$, $k = 8$)

Intervals	Observed frequency f_{i}	Expected frequency E_{i}
325–400	7	6.10
400–475	9	7.70
475–550	17	12.60
550–625	12	16.80
625–700	18	18.10
700–725	11	15.90
725–800	12	11.40
>800	14	11.40
	N = 100	

$$\chi^2_{5df}(0.90) = 9.24$$

Since $S = 5.404 < 9.24$, we would not reject the hypothesis of normality.

3.7.2 Kolmogorov–Smirnov (KS) Test

Both the χ^2 and KS tests are non-parameters. However, the χ^2 assumes large sample normality of the observed frequency about its mean while the KS test only assumes a continuous distribution. Suppose the hypothesis is that the sample comes from a specified population distribution with c.d.f. $F_0(x)$. The test statistic compares the observed (empirical) distribution of the sample $F_n(x)$, to $F_0(x)$ and considers the maximum absolute deviation between the two cdfs ($F_n(x)$ and F_0(x)). Let $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \ldots \leq X_{(n)}$ denote the ordered sample values. Define the observed distribution function, $F_n(x)$, as follows:

$$F_n(x) = \begin{cases} 0 \text{ for } x \leq x_{(1)} \\ \dfrac{i}{n} \text{ for } x_{(i)} < x \leq x_{(i+1)} \\ 1 \text{ for } x > x_{(n)} \end{cases} \tag{3.62}$$

Assume the testing hypothesis

$$H_0 : F(x) = F_0(x)$$

where $F_0(x)$ *is* a given continuous distribution and $F(x)$ is an unknown distribution. The KS test statistic can be computed as follows:

$$D_n = \max_{-\infty < x < \infty} |F_n(x) - F_0(x)|. \tag{3.63}$$

Since $F_0(x)$ is a continuous increasing function, we can evaluate $|F_n(x)–F_0(x)|$ for each n. Table A.4 in Appendix A (Smirnov 1948) gives certain critical values $d_{n,\alpha}$ of the distribution of D_n, which is the maximum absolute difference between sample and population cumulative distributions, for various sample sizes n and a is the level of significance. For example, the critical value $d_{n,\alpha}$ for $n = 10$ at a 0.05 level of significance is 0.409. This means that in 5 percent of random samples of size 10, the maximum absolute deviation between the sample observed cumulative distribution and the population cumulative distribtuion will be at least 0.409.

If $D_n \leq d_{n,\alpha}$ then we would not reject the hypothesis distribution H_o. If $D_n > d_{n,\alpha}$, then we reject the hypothetical distribution. The value $d_{n,\alpha}$ can be found in Table A.4 in Appendix A. This test implies that

$$P\{D_n > d_{n,\alpha}\} = \alpha.$$

Substituting for D_n and rearranging gives a 100(1–α)% confidence level for $F(x)$ as follows:

$$\hat{F}(x) - d_{n,\alpha} \leq F(x) \leq \hat{F}(x) + d_{n,\alpha}$$

or, equivalently, that

$$P\{\hat{F}(x) - d_{n,\alpha} \leq F(x) \leq \hat{F}(x) + d_{n,\alpha}\} = 1 - \alpha.$$

Example 3.26

(*Continued from Example* 3.25) Can the data given in Table 3.4 be represented by the Poisson distribution with a significance level of α using KS test?

To obtain the calculation in Table 3.6:

$F_n(x)$: is the cumulative observed frequency/33.

$F_0(x)$: is the estimated from cumulative $P(d)$ (see table below).

where

$P(d) = \frac{e^{-\lambda}\,\lambda^{\mathrm{d}}}{\mathrm{d}!}$ for d = 0, 1, …, 4

and $\lambda = 1.7$. Thus,

Table 3.6 KS test calculations ($N = 33$)

Number of defective	Observed frequency	Cum. observed frequency	$F_n(x)$	$F_0(x)$	$\lvert F_n(x) - F_0(x)\rvert$
0	6	6	0.182	0.183	0.001
1	8	14	0.424	0.494	0.07
2	12	26	0.788	0.758	0.03
3	4	30	0.909	0.907	0.002
4 or more	3	33	1.000	1.000	0.000

Table 3.7 KS test

Number of defective	Observed frequency f_i	$P(d)$	Cumulative$P(d)$
0	6	0.183	0.183
1	8	0.311	0.494
2	12	0.264	0.758
3	4	0.149	0.907
4 or more	3	0.093	1.000
Total	33	1.000	

$$D_n = \max_{-\infty<x<\infty} |F_n(x) - F_0(x)| = 0.070$$

From Table A.4 in Appendix A, here $N = 33$ and $\alpha = 10\%$ then we obtain $d_{n,\alpha} = 0.21$. Since $0.070 < 0.21$ ($D_n \leq d_{n,\alpha}$) therefore, the hypothesis of the Poisson model is accepted.

Example 3.27

Determine whether the following failure data (in days) of a system:

10.50, 1.75, 6.10, 1.30, 15.00, 8.20, 0.50, 20.50, 11.05, 4.60.

be represented as a sample from an exponential population distribution with constant rate $\lambda = 0.20$ failures per day at the ($\alpha =$) 5% level of the significance using KS test?

Solution: Under the hypothesis that failure times are exponential distribution, so the theoretical pdf and cdf are given by:

$$f(x) = 0.2e^{-0.2x} \text{ for } x > 0$$

and

$$F(x) = 1 - e^{-0.2x}$$

respectively. From Table 3.8, the maximum difference D_n is 0.2061. From Table A.4 in Appendix A, here n = 10 and $\alpha = 5\%$ then we obtain the critical value $d_{n,\alpha} = 0.409$. Since $D_n \leq d_{n,\alpha}$ therefore, the null hypothesis, that failure times are exponentially distributed with constant rate $\lambda = 0.20$, cannot be rejected at the 5% level of significance.

Table 3.8 The observed and theoretical cdf values

Failure time x	$F_n(x)$	$F_0(x)$	$\lvert F_n(x)–F_0(x)\rvert$
0.50	0.10	0.0952	0.0048
1.30	0.20	0.2289	0.0289
1.75	0.30	0.2953	0.0047
4.60	0.40	0.6015	0.2015
6.10	0.50	0.7048	0.2048
8.20	0.60	0.8061	0.2061
10.50	0.70	0.8776	0.1776
11.05	0.80	0.8903	0.0903
15.00	0.90	0.9503	0.0503
20.50	1.00	0.9835	0.0165

3.8 Test for Independence

The n elements of a sample may be classified according to two different criteria. It is of interest to know whether the two methods of classification are statistically independent. Assume that the first method of classification has r levels and the second method of classification has c levels. Let O_{ij} be the observed frequency for level i of the first classification method and level j of the second classification method as shown in Table 3.9 for $i = 1,2,\ldots,r$ and $j = 1,2,\ldots,c$.

We are interested in testing the hypothesis that the row and column methods of classification are independent. If we reject this hypothesis, we conclude there is some interaction between the two criteria of classification. The test statistic

$$\lambda^2_{\text{vahe}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \sim \lambda^2_{(r-1)(c-1)} \tag{3.64}$$

and we would reject the hypothesis of independence if $\lambda^2_{\text{value}} > \lambda^2_{\alpha,(r-1)(c-1)}$ where E_{ij}, the expected number in each cell, is

$$E_{ij} = \frac{1}{n}\left(\sum_{i=1}^{r} O_{ij}\right)\left(\sum_{j=1}^{c} O_{ij}\right).$$

For example,

Table 3.9 Observed frequency for level i and level j of classification

	1	2	3 …	c	Total
1	O_{11}	O_{12}	O_{13}	O_{1c}	$\sum_{j=1}^{c} O_{1j}$
2	O_{21}	O_{22}	O_{23}	O_{2c}	$\sum_{j=1}^{c} O_{2j}$
3	O_{31}	O_{32}	O_{33}	O_{3c}	$\sum_{j=1}^{c} O_{3j}$
…					
r	O_{r1}	O_{r2}	O_{r3}	O_{rc}	$\sum_{j=1}^{c} O_{rj}$
Total	$\sum_{i=1}^{r} O_{i1}$	$\sum_{i=1}^{r} O_{i2}$	$\sum_{i=1}^{r} O_{i3}$	$\sum_{i=1}^{r} O_{ic}$	n

Table 3.10 A random sample of 395 people

		Variable B				
VarA/VarB	Observed	15–20	21–30	31–40	41–65	Total
	Yes	60	54	46	41	201
Variable A	No	40	44	53	57	194
	Total	100	98	99	98	395

$$E_{21} = \frac{1}{n}\left(\sum_{i=1}^{r} O_{i1}\right)\left(\sum_{j=1}^{c} O_{2j}\right)$$
$$E_{32} = \frac{1}{n}\left(\sum_{i=1}^{r} O_{i2}\right)\left(\sum_{j=1}^{c} O_{3j}\right).$$

Example 3.28
A random sample of 395 people were surveyed. Each person was asked their interest in riding a bicycle (Variable A) and their age (Variable B). The data that resulted from the survey is summarized in Table 3.10.

Is there evidence to conclude, at the 0.05 level, that the desire to ride a bicycle depends on age? In other words, is age independent of the desire to ride a bicycle?

Solution: Here $r = 2$ and $c = 4$. The expected number in each cell can be obtained as shown in Table 3.11:

$$E_{11} = \frac{1}{395}\left(\sum_{i=1}^{2} O_{i1}\right)\left(\sum_{j=1}^{4} O_{1j}\right) = \frac{1}{395}(100)(201) = 50.8861$$
$$E_{12} = \frac{1}{395}\left(\sum_{i=1}^{2} O_{i2}\right)\left(\sum_{j=1}^{4} O_{1j}\right) = \frac{1}{395}(98)(201) = 49.8684$$
$$E_{13} = \frac{1}{395}(99)(201) = 50.3772$$
$$\cdots$$
$$E_{23} = \frac{1}{395}(99)(194) = 48.6228$$
$$E_{24} = \frac{1}{395}(98)(194) = 48.1316$$

The test statistic value

$$\lambda^2_{value} = \sum_{i=1}^{2}\sum_{j=1}^{4} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

Table 3.11 The expected values

		Variable B				
VarA/VarB	Expected	15–20	21–30	31–40	41–65	Total
	Yes	50.8861	49.8684	50.3772	49.8684	201
Variable A	No	49.1139	48.1316	48.6228	48.1316	194
	Total	100	98	99	98	395

$$= \frac{(60 - 50.8861)^2}{50.8861} + \frac{(54 - 49.8684)^2}{49.8684} + \ldots + \frac{(57 - 48.1316)^2}{48.1316}$$
$$= 8.0062$$

Here $\alpha = 0.05$. So $\chi^2_{0.05,(2-1)(4-1)} = \chi^2_{0.05,3} = 7.815$. Since $\chi^2_{\text{value}} = 8.0062 > 7.815$ we, therefore, reject the null hypothesis. That is, there is sufficient evidence, at the 0.05 level, to conclude that the desire to ride a bicycle depends on age.

3.9 Statistical Trend Tests of Homogeneous Poisson Process

Let us assume that repairable units observe from time $t = 0$, with successive failure times denoted by t_1, t_2, ... An equivalent representation of the failure process is in terms of the counting process $\{N(t),\ t \geq 0\}$, where $N(t)$ equals the number of failures in $(0,\ t]$. Repairable units are those units which can be restored to fully satisfactory performance by a method other than replacement of entire system (Ascher and Feingold 1984). It is assumed that simultaneous failures are not possible and the repair times are negligible compared to times between failures. In other word, it is assumed that repair times equal 0. The observed failure times during the interval (0,T] for a specific unit are denoted by $t_1,\ t_2,\ \ldots t_{N(T)}$.

Trend analysis is a common statistical method used to investigate the changes in a system device or unit over time. Based on the historical failure data of a unit or a group of similar units, one can test to determine whether the recurrent failures exhibit an increasing or decreasing trend subject to statistical tests. In other words, there are several methods to test whether the observed trend based on the failure data is statistically significant or there is no trend. This implies that, for the case there is no trend, the failure rate is constant or the failure process is homogeneous Poisson process. A number of tests such as Laplace Test and Military Handbook Test have been developed for testing the following null hypotheses (see Ascher and Feingold 1984):

H_0: No trend (Failure times observed are independently and identically distributed (iid), homogeneous Poisson process).

Against the alternative hypothesis.

H_1: Monotonic trend, i.e. the process is nonhomogeneous Poisson process (*NHPP*).

By rejection of null hypothesis, (case H_1) failure rate is not constant and the cumulative no of fault vs. time have concave or convex shape, respectively. Let t_i denote the running time of repairable item at the i*th* failure, $i = 1,\ldots,n$ and $N(t_i)$ be the total number of failures observed till time t_n(failure truncated case) or the observation is terminated at time T censored time for time truncated case which is discusses in the following subsections. Following we discuss the Laplace trend test and Military Handbook test. These tests can be used to detect existence of trends in the data set of successive failure times.

3.9.1 Laplace Trend Test

The Laplace trend test can be used to detect existence of trends in the failure data of successive event times. Depend on the data type whether is time truncated or failure truncated, we can calculate Laplace trend index as follows:

Time Truncated Test: Suppose that observation is terminated at time T and $N(t_n)$ failures are recorded at time $t_1 < t_2 < \ldots < t_n < T$. The test statistic U of Laplace trend test in this case is given by

$$U = \sqrt{12.N(t_n)}\left(\frac{\sum_1^n t_i}{T.N(t_n)} - 0.5\right) \tag{3.65}$$

Failure Truncated Test: In the cases where the observation terminates at a failure event, say t_n, instead of T, the test statistics for failure truncated data is given by

$$U = \sqrt{12\ N(t_{n-1})}\left(\frac{\sum_1^{n-1} t_i}{t_n\ N(t_{n-1})} - 0.5\right). \tag{3.66}$$

In both test cases, the test statistic U is approximately standard normally $N(0, 1)$ distributed when the null hypothesis H_0 is true.

At significance level α of 5% for example, the lower and upper bound of confidence interval are -1.96 and $+\ 1.96$, respectively. If U is less than lower bound or greater than upper bound then there is a trend in the data. In these cases, we can reject the null hypothesis of trend free at 5% and accept the alternative hypothesis. In the first case, there is reliability growth (U is less than lower bound) and in second case (U is greater than upper bound) there is reliability deterioration. Hence, in these cases, one need to be some probability distributions that can fit to the failure data set, other than exponential distribution.

In general, at significance level α, the rejection criteria of the null hypothesis is given by.

$$\text{Reject } H_0 : U < -z_{\alpha/2} \quad \text{or} \quad U > z_{\alpha/2} \tag{3.67}$$

where z_α is $(1-\alpha)$th quantile of the standard normal distribution.

3.9.2 Military Handbook Test

This test is based on the assumption that a failure intensity of $r(t) = \lambda\beta\, t^{\beta-1}$ is appropriate (MIL-HDBK-189 1981). When $\beta = 1$, the failure intensity reduces to $r(t) = \lambda$, which means the failure process follows a HPP. Hence the test determines whether an estimate of β is significantly different from 1 or not. The null hypothesis test in this case is $\beta = 1$ i.e. no trend in data (HPP) vs. alternative hypothesis $\beta \neq 1$, there is trend in the failure history (NHPP):

Time Truncated Test: During a test period T suppose that n failure are recorded at time $t_1 < t_2 < \ldots < t_n < T$.The mathematical test in this case is given by:

$$S = 2\sum_{i=1}^{n} \ln \frac{T}{t_i}. \tag{3.68}$$

According to (MIL-HDBK-189, 1981), S has chi-square distribution with $2n$ degree of freedom. In this case, the rejection criteria is given by:

$$\text{Reject } H_0 \text{ when } S < \chi^2_{2n,\,1-\alpha/2} \quad \text{or} \quad S > \chi^2_{2n,\,\alpha/2}. \tag{3.69}$$

In other words, the hypothesis (H_0) of trend free is rejected for the values beyond the interval boundaries. Low values of S correspond to deteriorating while large values of S correspond to improving.

Failure Truncated Test: If the data are failure truncated at t_n instead of time truncated at T, S has chi-square distribution with $2(n-1)$ degree of freedom when the null hypothesis H_0 is true. The test statistic S of Military handbook test in this case is given by:

$$S = 2\sum_{i=1}^{n-1} \ln\left(\frac{t_n}{t_i}\right). \tag{3.70}$$

Here, the rejection criterion is given by:

$$\text{Reject } H_0 \text{ when } S < \chi^2_{2(n-1),\,1-\alpha/2} \quad \text{or} \quad S > \chi^2_{2(n-1),\,\alpha/2}. \tag{3.71}$$

The hypothesis of HPP (H_0) is rejected for the values beyond the following interval:

$$\left[\chi^2_{2(n-1),1-\alpha/2},\, \chi^2_{2(n-1),\alpha/2}\right].$$

3.10 Least Squared Estimation

A problem of curve fitting, which is unrelated to normal regression theory and MLE estimates of coefficients but uses identical formulas, is called the method of least squares. This method is based on minimizing the sum of the squared distance from the best fit line and the actual data points. It just so happens that finding the MLEs for the coefficients of the regression line also involves these sums of squared distances.

Normal Linear Regression

Regression considers the distributions of one variable when another is held fixed at each of several levels. In the bivariate normal case, consider the distribution of X as a function of given values of Z where $X = \alpha + \beta Z$. Consider a sample of n observations (x_i, z_i), we can obtain the likelihood and its natural log for the normal distribution as follows:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{2\pi^{\frac{n}{2}}}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\alpha-\beta z_i)^2}$$

$$\ln L = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha - \beta z_i)^2. \tag{3.72}$$

Taking the partial derivatives of $\ln L$ with respect to α and β, we have

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n}(x_i - \alpha - \beta z_i)^2 = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} z_i(x_i - \alpha - \beta z_i) = 0.$$

The solution of the simultaneous equations is

$$\hat{\alpha} = \bar{X} - \beta\bar{Z}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Z_i - \bar{Z})}{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}. \tag{3.73}$$

Least Squared Straight Line Fit

Assume there is a linear relationship between X and $E(Y/x)$, that is, $E(Y/x) = a + bx$. Given a set of data, we want to estimate the coefficients a and b that minimize the sum of the squares. Suppose the desired polynomial, $p(x)$, is written as

$$\sum_{i=0}^{m} a_i x^i$$

where a_0, a_1,…, a_m are to be determined. The method of least squares chooses as the solutions those coefficients minimizing the sum of the squares of the vertical distances from the data points to the presumed polynomial. This means that the polynomial termed "best" is the one whose coefficients minimize the function L, where

$$L = \sum_{i=1}^{n} [y_i - p(x_i)]^2.$$

Here, we will treat only the linear case, where $X = \alpha + \beta Z$. The procedure for higher-order polynomials is identical, although the computations become much more tedious. Assume a straight line of the form $X = \alpha + \beta Z$. For each observation (x_i, z_i): $X_i = \alpha + \beta Z_i$, let

$$Q = \sum_{i=1}^{n} (x_i - \alpha - \beta z_i)^2.$$

We wish to find α and β estimates such as to minimize Q. Taking the partial differentials, we obtain

$$\frac{\partial Q}{\partial \alpha} = -2 \sum_{i=1}^{n} (x_i - \alpha - \beta z_i) = 0$$
$$\frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^{n} z_i (x_i - \alpha - \beta z_i) = 0.$$

Note that the above are the same as the MLE equations for normal linear regression. Therefore, we obtain the following results:

$$\hat{\alpha} = \bar{x} - \beta \bar{z}$$
$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}. \tag{3.74}$$

The above gives an example of least squares applied to a linear case. It follows the same pattern for higher-order curves with solutions of 3, 4, and so on, from the linear systems of equations.

3.11 Interval Estimation

A point estimate is sometimes inadequate in providing an estimate of an unknown parameter since it rarely coincides with the true value of the parameter. An alternative

way is to obtain a confidence interval estimation of the form $[\theta_L, \theta_U]$ where θ_L is the lower bound and θ_U is the upper bound. A confidence interval (CI) is an interval estimate of a population parameter that also specifies the likelihood that the interval contains the true population parameter. This probability is called the level of confidence, denoted by $(1-\alpha)$, and is usually expressed as a percentage.

Point estimates can become more useful if some measure of their error can be developed, i.e., some sort of tolerance on their high and low values could be developed. Thus, if an interval estimator is $[\theta_L, \theta_U]$ with a given probability $(1-\alpha)$, then θ_L and θ_U will be called $100(1-\alpha)\%$ confidence limits for the given parameter θ and the interval between them is a $100(1-\alpha)\%$ confidence interval and $(1-\alpha)$ is also called the confidence coefficient.

3.11.1 Confidence Intervals for the Normal Parameters

The one-dimensional normal distribution has two parameters: mean μ and variance σ^2. The simultaneous employment of both parameters in a confidence statement concerning percentages of the population will be discussed in the next section on tolerance limits. Hence, individual confidence statements about μ and σ^2 will be discussed here.

Confidence Limits for the Mean μ with Known σ^2.

It is easy to show that the statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal distribution where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Hence, a $100(1-\alpha)\%$ confidence interval for the mean μ is given by

$$P\left[\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha \tag{3.75}$$

In other words,

$$\mu_L = \bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \mu_U = \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

For the normal distribution, an empirical rule relates the standard deviation to the proportion of the observed values of the variable in a data set that lie in an interval around the sample mean can be approximately expressed as follows:

- 68% of the values lie within: $\bar{x} \pm s_x$
- 95% of the values lie within: $\bar{x} \pm 2s_x$
- 99.7% of the values lie within: $\bar{x} \pm 3s_x$

Example 3.29
Draw a sample of size 4 from a normal distribution with known variance = 9, say $x_1 = 2$, $x_2 = 3$, $x_3 = 5$, $x_4 = 2$. Determine the location of the true mean (μ). The sample mean can be calculated as.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+3+5+2}{4} = 3$$

Assuming that $\alpha = 0.05$ and from the standard normal distribution (Table A.1 in Appendix A), we obtain

$$P\left[3 - 1.96\frac{3}{\sqrt{4}} < \mu < 3 + 1.96\frac{3}{\sqrt{4}}\right] = 0.95$$
$$P[0.06 < \mu < 5.94] = 0.95$$

This example shows that there is a 95% probability that the true mean is somewhere between 0.06 and 5.94. Now, μ is a fixed parameter and does not vary, so how do we interpret the probability? If the samples of size 4 are repeatedly drawn, a different set of limits would be constructed each time. With this as the case, the interval becomes the random variable and the interpretation is that, for 95% of the time, the interval so constructed will contain the true (fixed) parameter.

Confidence Limits for the Mean μ with Unknown σ^2.

Let

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{3.76}$$

It can be shown that the statistic

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

has a t distribution with $(n-1)$ degrees of freedom (see Table A.2 in Appendix A). Thus, for a given sample mean and sample standard deviation, we obtain

$$P\left[|T| < t_{\frac{\alpha}{2},n-1}\right] = 1 - \alpha$$

Hence, a $100(1-\alpha)\%$ confidence interval for the mean μ is given by

$$p\left[\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right] = 1 - \alpha \tag{3.77}$$

Example 3.30
A problem on the variability of a new product was encountered. An experiment was run using a sample of size $n = 25$; the sample mean was found to be $\bar{X} = 50$ and the variance $\sigma^2 = 16$. From Table A.2 in Appendix A, $t_{\frac{\alpha}{2},n-1} = t_{.975,24} = 2.064$. A 95% confidence limit for μ is given by.

$$P\left[50 - 2.064\sqrt{\frac{16}{25}} < \mu < 50 + 2.064\sqrt{\frac{16}{25}}\right] = 0.95$$

$$P[48.349 < \mu < 51.651] = 0.95$$

Note that, for one-sided limits, choose t_{α}, or $t_{1-\alpha}$.

Example 3.31
Consider a normal distribution with unknown mean μ and unknown standard deviation σ. Suppose we draw a random sample of size n = 16 from this population, and the sample values are:

16.16	9.33	12.96	11.49
12.31	8.93	6.02	10.66
7.75	15.55	3.58	11.34
11.38	6.53	9.75	9.47

Compute the confidence interval for the mean μ and at confidence level $1 - \alpha = 0.95$.

Solution: A C.I. for μ at confidence level 0.95 when σ is unknown is obtained from the corresponding t-test. The C.I is

$$\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}$$

The sample mean and sample variance, respectively, are:

$$\bar{X} = 10.20 \; and \; S^2 = 10.977$$

The sample standard deviation is: $S = 3.313$. Also, $t_{1-\frac{\alpha}{2},n-1} = t_{0.975,15} = 2.131$.

The 95% C.I. for μ is

$$10.20 - 2.131\frac{3.313}{\sqrt{16}} \leq \mu \leq 10.20 + 2.131\frac{3.313}{\sqrt{16}}$$

or

$$8.435 \leq \mu \leq 11.965$$

Confidence Limits on σ^2.

Note that $n\frac{\widehat{\sigma^2}}{\sigma^2}$ has a χ^2 distribution with $(n-1)$ degrees of freedom. Correcting for the bias in $\overset{\wedge}{\sigma^2}$, then $(n-1)\,\overset{\wedge}{\sigma^2}/\sigma^2$ has this same distribution. Hence,

$$P\left[\chi^2_{\frac{\alpha}{2},n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\frac{\alpha}{2},n-1}\right] = 1-\alpha$$

or

$$P\left[\frac{\sum(x_i-\bar{x})^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{\sum(x_i-\bar{x})^2}{\chi^2_{\frac{\alpha}{2},n-1}}\right] = 1-\alpha. \tag{3.78}$$

Similarly, for one-sided limits, choose $\chi^2(\alpha)$ or $\chi^2(1-\alpha)$.

3.11.2 Confidence Intervals for the Exponential Parameters

The pdf and cdf of the exponential distribution are given as

$$f(x) = \lambda e^{-\lambda x} \qquad x > 0, \lambda > 0$$

and

$$F(x) = 1 - e^{-\lambda x}$$

respectively. From Eq. (3.22), it was shown that the distribution of a function of the estimate

$$\overset{\wedge}{\lambda} = \frac{r}{\sum_{i=1}^{r} x_i + (n-r)x_r} \tag{3.79}$$

derived from a test of n identical components with common exponential failure density (failure rate λ), whose testing was stopped after the rth failure, was chi-squared (X^2), i.e.,

$$2r\frac{\lambda}{\hat{\lambda}} = 2\lambda T \quad (\chi^2 \text{distribution with } 2r \text{ degrees of freedom})$$

where T is the total accrued time on all units. Knowing the distribution of $2\lambda T$ allows us to obtain the confidence limits on the parameter as follows:

$$P\left[\chi^2_{1-\frac{\alpha}{2},2r} < 2\lambda T < \chi^2_{\frac{\alpha}{2},2r}\right] = 1-\alpha$$

or, equivalently, that

$$P\left[\frac{\chi^2_{1-\frac{\alpha}{2},2r}}{2T} < \lambda < \frac{\chi^2_{\frac{\alpha}{2},2r}}{2T}\right] = 1-\alpha$$

This means that in $(1-\alpha)\%$ of samples with a given size n, the random interval

$$\left(\frac{\chi^2_{1-\frac{\alpha}{2},2r}}{2T}, \frac{\chi^2_{\frac{\alpha}{2},2r}}{2T}\right) \tag{3.80}$$

will contain the population of constant failure rate. In terms of $\theta = 1/\lambda$ or the mean time between failure (MTBF), the above confidence limits change to

$$P\left[\frac{2T}{\chi^2_{\frac{\alpha}{2},2r}} < \theta < \frac{2T}{\chi^2_{1-\frac{\alpha}{2},2r}}\right] = 1-\alpha.$$

If testing is stopped at a fixed time rather than a fixed number of failures, the number of degrees of freedom in the lower limit increases by two. Table 3.12 shows

Table 3.12 Confidence limits for θ

Confidence limits	Fixed number of failures	Fixed time
One-sided lower limit	$\frac{2T}{\chi^2_{\alpha,2r}}$	$\frac{2T}{\chi^2_{\alpha,2r+2}}$
One-sided upper limit	$\frac{2T}{\chi^2_{1-\alpha,2r}}$	$\frac{2T}{\chi^2_{1-\alpha,2r}}$
Two-sided limits	$\frac{2T}{\chi^2_{\alpha/2,2r}}, \frac{2T}{\chi^2_{1-\alpha/2,2r}}$	$\frac{2T}{\chi^2_{\alpha/2,2r+2}}, \frac{2T}{\chi^2_{1-\alpha/2,2r}}$

the confidence limits for θ, the mean of an exponential density.

Example 3.32
(*Two-sided*) From the goodness of fit example, $T = 22{,}850$, testing stopped after $r = 60$ failures. We can obtain $\hat{\lambda} = 0.00263$ and $\hat{\theta} = 380.833$. Assu-ming that $\alpha = 0.1$, then, from the above formula, we obtain.

$$P\left[\frac{2T}{\chi^2_{0.05,120}} < \theta < \frac{2T}{\chi^2_{0.95,120}}\right] = 0.9$$
$$P\left[\frac{45,700}{146.568} < \theta < \frac{45,700}{95.703}\right] = 0.9$$
$$P[311.80 < \theta < 477.52] = 0.9$$

Example 3.33
(*One-sided lower*) Assuming that testing stopped after 1,000 h with four failures, then.

$$P\left[\frac{2T}{\chi^2_{0.10,10}} < \theta\right] = 0.9$$
$$P\left[\frac{2,000}{15.987} < \theta\right] = 0.9$$
$$P[125.1 < \theta] = 0.9$$

3.11.3 Confidence Intervals for the Binomial Parameters

Consider a sequence of n Bernoulli trials with k successes and $(n - k)$ failures. We now determine one-sided upper and lower and two-sided limits on the parameter p, the probability of success. For the lower limit, the binomial sum is set up such that the chance probability of k or more successes with a true p as low as p_L is only $\alpha/2$. This means the probability of k or more successes with a true p higher than p_L is $\left(1 - \frac{\alpha}{2}\right)$:

$$\sum_{i=k}^{n} \binom{n}{i} p_L^i (1 - p_L)^{n-i} = \frac{\alpha}{2}$$

Similarly, the binomial sum for the upper limit is

$$\sum_{i=k}^{n} \binom{n}{i} p_U^i (1-p_U)^{n-i} = 1 - \frac{\alpha}{2}$$

or, equivalently, that

$$\sum_{i=0}^{k-1} \binom{n}{i} p_U^i (1-p_U)^{n-i} = \frac{\alpha}{2}$$

Solving for p_L and p_U in the above equations,

$$P[p_L < p < p_U] = 1 - \alpha$$

For the case of one-sided limits, merely change $\alpha/2$ to α.

Example 3.34
Given $n = 100$ with 25 successes, and 75 failures, an 80% two-sided confidence limits on p can be obtained as follows:

$$\sum_{i=25}^{100} \binom{100}{i} p_L^i (1-p_L)^{100-i} = 0.10$$
$$\sum_{i=0}^{24} \binom{100}{i} p_U^i (1-p_U)^{100-i} = 0.10$$

Solving the above two equations simultaneously, we obtain

$$p_L \approx 0.194 \text{ and } p_U \approx 0.313$$
$$P[0.194 < p < 0.313] = 0.80$$

Example 3.35
Continuing with Example 3.34, find an 80% one-sided confidence limit on p.

Solution: We now can set the top equation to 0.20 and solve for p_L. It is easy to obtain $p_L = 0.211$ and $P[p > 0.211] = 0.80$. Let us define $\bar{p} = k/n$, the number of successes divided by the number of trials. For large values of n and if $np > 5$ and $n(1-p) > 5$, and from the central limit theorem (Feller 1966, 1968), the statistic

$$Z = \frac{(\bar{p} - p)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \tag{3.81}$$

approximates to the standard normal distribution. Hence,

$$P[-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}] = 1 - \alpha$$

Then

$$P\left[\bar{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} < p < \bar{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right] = 1 - \alpha \tag{3.82}$$

Example 3.36
Given $n = 900$, $k = 180$, and $\alpha = 0.05$. Then we obtain $p = 180/900 = 0.2$ and

$$P\left[0.2 - 1.96\sqrt{\frac{0.2(0.8)}{900}} < p < 0.2 + 1.96\sqrt{\frac{0.2(0.8)}{900}}\right] = 0.95$$
$$P[0.174 < p < 0.226] = 0.95$$

3.11.4 Confidence Intervals for the Poisson Parameters

Limits for the Poisson parameters are completely analogous to the binomial except that the sample space is infinite instead of finite. The lower and upper limits can be solved simultaneously in the following equations:

$$\sum_{i=k}^{\infty} \frac{\lambda_L^i e^{-\lambda_L}}{i!} = \frac{\alpha}{2} \quad \text{and} \quad \sum_{i=k}^{\infty} \frac{\lambda_U^i e^{-\lambda_U}}{i!} = 1 - \frac{\alpha}{2} \tag{3.83}$$

or, equivalently

$$\sum_{i=k}^{\infty} \frac{\lambda_L^i e^{-\lambda_L}}{i!} = \frac{\alpha}{2} \quad \text{and} \quad \sum_{i=0}^{k-1} \frac{\lambda_U^i e^{-\lambda_U}}{i!} = \frac{\alpha}{2}.$$

Example 3.37
One thousand article lots are inspected resulting in an average of 10 defects per lot. Find 90% limits on the average number of defects per 1000 article lots. Assume $\alpha = 0.1$,

$$\sum_{i=10}^{\infty} \frac{\lambda_L^i e^{-\lambda_L}}{i!} = 0.05 \quad \text{and} \quad \sum_{i=0}^{9} \frac{\lambda_U^i e^{-\lambda_U}}{i!} = 0.05.$$

Solving the above two equations simulteneously for λ_L and λ_U, we obtain.
$P\ [5.45 < \lambda < 16.95] = 0.90$.
The one-sided limits are constructed similarly to the case for binomial limits.

3.12 Non-Parametric Tolerance Limits

Non-parametric tolerance limits are based on the smallest and largest observation in the sample, designated as X_S and X_L, respectively. Due to their non-parametric nature, these limits are quite insensitive and to gain precision proportional to the parametric methods requires much larger samples. An interesting question here is to determine the sample size required to include at least $100(1-\alpha)\%$ of the population between X_S and X_L with given probability γ.

For two-sided tolerance limits, if $(1-\alpha)$ is the minimum proportion of the population contained between the largest observation X_L and smallest observation X_S with confidence $(1-\gamma)$, then it can be shown that

$$n(1-\alpha)^{n-1} - (n-1)(1-\alpha)^n = \gamma. \tag{3.84}$$

Therefore, the number of observations required is given by

$$n = \left\lfloor \frac{(2-\alpha)}{4\alpha} \chi^2_{1-\gamma,4} + \frac{1}{2} \right\rfloor + 1 \tag{3.85}$$

where a value of $\chi^2_{1-\gamma,4}$ is given in Table A.3 of Appendix A.

Example 3.38
Determine the tolerance limits which include at least 90% of the population with probability 0.95. Here,

$\alpha = 0.1$, $\gamma = 0.95$ and ${\chi^2}_{0.05,4} = 9.488$.

and therefore, a sample of size

$$n = \left\lfloor \frac{(2-0.1)}{4(0.1)}(9.488) + \frac{1}{2} \right\rfloor + 1 = 46$$

is required. For a one-sided tolerance limit, the number of observations required is given by

$$n = \left\lfloor \frac{\log(1-\gamma)}{\log(1-\alpha)} \right\rfloor + 1. \tag{3.86}$$

Example 3.39
As in Example 3.38, we wish to find a lower tolerance limit, that is, the number of observations required so that the probability is 0.95 that at least 90% of the population will exceed X_S is given by.

$$n = \left\lfloor \frac{\log(1-0.95)}{\log(1-0.1)} \right\rfloor + 1 = 30.$$

One can easily generate a table containing the sample size required to include a given percentage of the population between X_S and X_L with given confidence, or sample size required to include a given percentage of the population above or below X_S or X_L, respectively.

3.13 Sequential Sampling

A sequential sampling scheme is one in which items are drawn one at a time and the results at any stage determine if sampling or testing should stop. Thus, any sampling procedure for which the number of observations is a random variable can be regarded as sequential sampling. Sequential tests derive their name from the fact that the sample size is not determined in advance, but allowed to "float" with a decision (accept, reject, or continue test) after each trial or data point.

In general, let us consider the hypothesis

$$H_0 : f(x) = f_0(x) \quad \text{vs} \quad H_1 : f(x) = f_1(x)$$

For an observation test, say X_1, if $X_1 \leq A$, then we will accept the testing hypothesis (H_0: $f(x) = f_0(x)$); if $X_1 \geq A$, then we will reject H_0 and accept H_1: $f(x) = f_1(x)$. Otherwise, we will continue to perform at least one more test. The interval $X_1 \leq A$ is called the acceptance region. The interval $X_1 \geq$ A is called the rejection or critical region (see Fig. 3.4).

A "good" test is one that makes the α and β errors as small as possible where

$$P\{\text{Type I error}\} = \text{P}\{\text{Reject H}_0 \mid \text{H}_0 \text{ is True}\} = \alpha$$
$$P\{\text{Type II error}\} = \text{P}\{\text{Accept H}_0 \mid \text{H}_0 \text{ is False}\} = \beta$$

Type I error, also known as a "*false positive*", the error of rejecting a null hypothesis H_o when it is actually true. In other words, this is the error of accepting an alternative hypothesis H_1 (the real hypothesis of interest) when the results can be attributed to chance. So the probability of making a type I error in a test with rejection region R is P(R | H_o is true).

Type II error, also known as a "*false negative*", the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. In other words, this is the error of failing to accept an alternative hypothesis when you don't have

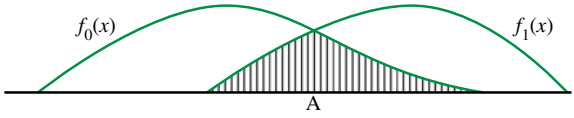

Fig. 3.4 A sequential sampling scheme

adequate power. So the probability of making a type II error in a test with rejection region R is 1-P(R | H_1 is true). The power of the test can be P(R | H_1 is true).

However, there is not much freedom to do this without increasing the sample size. The common procedure is to fix the β error and then choose a critical region to minimize the error or maximize the "power" (power $= 1-\beta$) of the test, or to choose the critical region so as to equalize the α and β errors to reasonable levels.

A criterion, similar to the MLE, for constructing tests is called the "sequential probability ratio", which is the ratio of the sample densities under H_1 over H_0. The sequential probability ratio is given by

$$\lambda_n = \frac{\prod_{i=1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)} > k \tag{3.87}$$

Here, $x_1, x_2, \ldots, x_n$ are n independent random observations and k is chosen to give the desired error.

Recall from the MLE discussion in Sect. 3.3 that $f_1(x_1)$, $f_1(x_2),\ldots, f_1(x_n)$ are maximized under H_1 when the parameter(s), e.g., $\theta = \theta_1$ and, similarly, $f_0(x_1)$, $f_0(x_2),\ldots, f_0(x_n)$ are maximized when $\theta = \theta_0$. Thus, the ratio will become large if the sample favors H_1 and will become small if the sample favors H_0. Therefore, the test will be called a sequential probability ratio test if we.

1. Stop sampling and reject H_0 as soon as $\lambda_n \geq A$.
2. Stop sampling and accept H_0 as soon as $\lambda_n \leq B$.
3. Continue sampling as long as $B < \lambda_n < A$, where $A > B$.

For example, the test will continue iff

$$\frac{\beta}{1-\alpha} < \lambda_n < \frac{1-\beta}{\alpha}$$

The choice of A and B with the above test, suggested by Wald (1947), can be determined as follows:

$$B = \frac{\beta}{1-\alpha} \text{ and } A = \frac{1-\beta}{\alpha}$$

The basis for α and β are therefore

$$P[\lambda_n \geq A | H_0] = \alpha$$
$$P[\lambda_n \leq B | H_1] = \beta$$

3.13.1 Exponential Distribution Case

Let

$$V(t) = \sum_{i=1}^{r} X_i + \sum_{j=1}^{n-r} t_j$$

where X_i are the times to failure and t_j are the times to test termination without failure. Thus, $V(t)$ is merely the total operating time accrued on both successful and unsuccessful units where the total number of units is n. The hypothesis to be tested is.

H_0: $\theta = \theta_0$ vs H_1: $\theta = \theta_1$.

For the failed items,

$$g(x_1, x_2, \ldots, x_r) = \left(\frac{1}{\theta}\right)^r e^{-\frac{\sum_{i=1}^{r} x_i}{\theta}}$$

For the non-failed items,

$$P(X_{r+1} > t_1, X_{r+2} > t_2, \ldots, X_n > t_{n-r}) = e^{-\frac{\sum_{j=1}^{n-r} t_j}{\theta}}$$

The joint density for the first r failures among n items, or likelihood function, is

$$f(x_1, x_2, \ldots, x_r, t_{r+1}, \ldots, t_n) = \frac{n!}{(n-r)!}\left(\frac{1}{\theta}\right)^r e^{-\frac{1}{\theta}\left(\sum_{i=1}^{r} x_i + \sum_{j=1}^{n-r} t_j\right)}$$

$$= \frac{n!}{(n-r)!}\left(\frac{1}{\theta}\right)^r e^{-\frac{V(t)}{\theta}}$$

and the sequential probability ratio is given by

$$\lambda_n = \frac{\prod_{i=1}^{n} f_1(x_i, \theta_1)}{\prod_{i=1}^{n} f_0(x_i, \theta_0)} = \frac{\frac{n!}{(n-r)!}\left(\frac{1}{\theta_1}\right)^r e^{-\frac{V(t)}{\theta_1}}}{\frac{n!}{(n-r)!}\left(\frac{1}{\theta_0}\right)^r e^{-\frac{V(t)}{\theta_0}}}$$

$$= \left(\frac{\theta_0}{\theta_1}\right)^r e^{-V(t)\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$$

Now, it has been shown that for sequential tests, the reject and accept limits, A and B, can be equated to simple functions of α and β. Thus, we obtain the following test procedures:

Continue test: $\frac{\beta}{1-\alpha} \equiv B < \lambda_n < A \equiv \frac{1-\beta}{\alpha}$.

Reject H_0: $\lambda_n > A \equiv \frac{1-\beta}{\alpha}$.

Accept H_0: $\lambda_n < B \equiv \frac{\beta}{1-\alpha}$.

Working with the continue test inequality, we now have

$$\frac{\beta}{1-\alpha} < \left(\frac{\theta_0}{\theta_1}\right)^r e^{-V(t)\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right]} < \frac{1-\beta}{\alpha}$$

Taking natural logs of the above inequality, we obtain

$$\ln\left(\frac{\beta}{1-\alpha}\right) < r\ln\left(\frac{\theta_0}{\theta_1}\right) - V(t)\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right] < \ln\left(\frac{1-\beta}{\alpha}\right)$$

The above inequality is linear in $V(t)$ and r, and therefore the rejection line $V(t)$, say $V_r(t)$, can be obtained by setting

$$r\ln\left(\frac{\theta_0}{\theta_1}\right) - V_r(t)\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right] = \ln\left(\frac{1-\beta}{\alpha}\right)$$

or, equivalently,

$$V_r(t) = \frac{r\ln\left(\frac{\theta_0}{\theta_1}\right)}{\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right]} - \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right]}$$

Similarly, the acceptance line $V(t)$), say $V_a(t)$, (see Fig. 3.5) can be obtained by setting

$$r\log\left(\frac{\theta_0}{\theta_1}\right) - V_a(t)\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right] = \log\left(\frac{\beta}{1-\alpha}\right)$$

This implies that

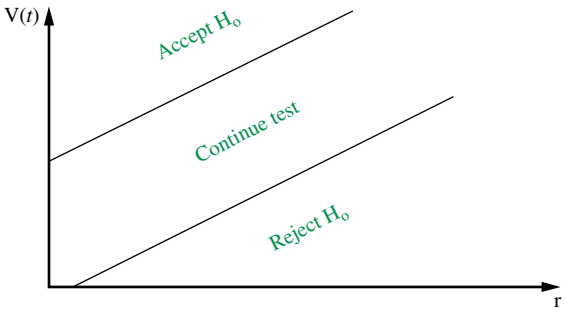

Fig 3.5 Test procedure

$$V_a(t) = \frac{r \ln\left(\frac{\theta_0}{\theta_1}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]} - \frac{\ln\left(\frac{\beta}{1-\alpha}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$$

Thus, continue to perform the test when

$$-h_1 + rs < V(t) < h_0 + rs$$

where

$h_0 = -\frac{\ln\left(\frac{\beta}{1-\alpha}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$, $s = \frac{\ln\left(\frac{\theta_0}{\theta_1}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$, $h_1 = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$, and

$$V(t) = \sum_{i=1}^{r} X_i + \sum_{j=1}^{n-r} t_j$$

Example 3.40
Given that H_0: $\theta = 500$ vs H_1: $\theta = 250$ and $\alpha = \beta = 0.1$. The acceptance and rejection lines are given by

$$V_a(t) = 346.6r + 1098.6$$

and

$$V_r(t) = 346.6r - 1098.6$$

respectively. Both are linear functions in terms of r, the number of first r failures in the test. For an exponential distribution

θ	$P(A)$
0	0
θ_1	β
$\log\left(\frac{\theta_0}{\theta_1}\right)\frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\left[\frac{1}{\theta_1} - \frac{1}{\theta_0}\right]}$	$\frac{1-\beta}{\log\left(\frac{1-\beta}{\alpha}\right) - \log\left(\frac{\beta}{1-\alpha}\right)}$
θ_1	$1-\alpha$
∞	1

From the information given in the above example, we can obtain

θ	$P(A)$
0	0
250	0.10
346.5	0.5
500	0.90
∞	1.0

Since there is no pre-assigned termination to a regular sequential test, it is customary to draw a curve called the "average sample number" (ASN). This curve shows the expected sample size as a function of the true parameter value. It is known that the test will be terminated with a finite observation. It should be noted that "on average", the sequential tests utilizes significantly smaller samples than fixed sample plans:

θ	$E(r) = ASN$
0	0
θ_1	$\frac{\theta_0\beta\log\left(\frac{\beta}{1-\alpha}\right)+(1-\beta)\log\left(\frac{1-\beta}{\alpha}\right)}{\left[\log\left(\frac{\theta_0}{\theta_1}\right)-\left(\frac{\theta_0-\theta_1}{\theta_1}\right)\right]\theta_1}$
$\frac{\log\left(\frac{\theta_0}{\theta_1}\right)}{\left[\frac{1}{\theta_1}-\frac{1}{\theta_0}\right]}$	$\frac{\log\left(\frac{1-\beta}{\alpha}\right)\log\left(\frac{\beta}{1-\alpha}\right)}{\left[\log\left(\frac{\theta_0}{\theta_1}\right)\right]^2}$
θ_0	$\frac{\left[(1-\alpha)\log\left(\frac{\beta}{1-\alpha}\right)+\alpha\log\left(\frac{1-\beta}{\alpha}\right)\right]\theta_1}{\left[\log\left(\frac{\theta_0}{\theta_1}\right)-\left(\frac{\theta_0-\theta_1}{\theta_1}\right)\right]\theta_0}$
∞	0

An approximate formula for $E(t)$, the expected time to reach decision, is

$$E(t) \equiv \theta \log\left(\frac{n}{n - E(r)}\right)$$

where *n is* the total number of units on test (assuming no replacement of failed units). If replacements are made, then

$$E(t) = \frac{\theta}{n} E(r)$$

Occasionally, it is desired to "truncate" a sequential plan such that, if no decision is made before a certain point, testing is stopped and a decision is made on the basis of data acquired up to that point (Pham 2000, page 251). There are a number of rules and theories on optimum truncation. In the reliability community, a $V(t)$ truncation point at $10\theta_0$ is often used to determine the $V(t)$ and r lines for truncation and the corresponding exact $\alpha = \beta$ errors (these will in general be larger for truncated tests than for the non-truncated). An approximate method draws the $V(t)$ truncation line to

the center of the continue test band and constructs the r truncation line perpendicular to that point.

3.13.1.1 Bernoulli Distribution Case

$$H_0: p = p_0 \text{ vs } H_1: p = p_1.$$

where p denotes the proportion defective units in the population, and α and β are pre-determined. Let us define the following random variable X:

$$X = \begin{cases} 0 & \text{if good} \\ 1 & \text{if defective} \end{cases}$$

where p denotes the proportion defective items in the population. The Bernoulli distribution is given by

$$P(x) = p^x(1-p)^{1-x} \quad \text{for } x = 0, 1$$

The sequential probability ratio is given by

$$\lambda_n = \frac{\prod_{i=1}^{n} P(x_i, p_1)}{\prod_{i=1}^{n} P(x_i, p_0)} = \frac{\prod_{i=1}^{n} p_1^{x_i}(1-p_1)^{1-x_i}}{\prod_{i=1}^{n} p_0^{x_i}(1-p_0)^{1-x_i}} = \left(\frac{p_1}{p_0}\right)^{\sum_{i=1}^{n} x_i} \left(\frac{1-p_1}{1-p_0}\right)^{n-\sum_{i=1}^{n} x_i}$$
$$= \left(\frac{p_1}{p_0}\right)^{d_n} \left(\frac{1-p_1}{1-p_0}\right)^{n-d_n}$$

where $d_n = \sum_{i=1}^{n} x_i$ is the cumulative number of defective units and n is the cumulative sample size. As we know, sampling process will continue as long as

$$\frac{\beta}{1-\alpha} < \lambda_n < \frac{1-\beta}{\alpha}$$

Taking the natural log, we have

$$\ln\left(\frac{\beta}{1-\alpha}\right) < \ln(\lambda_n) < \ln\left(\frac{1-\beta}{\alpha}\right)$$

or equivalently that,

$$\ln\left(\frac{\beta}{1-\alpha}\right) < d_n \ln\left(\frac{p_1}{p_0}\right) + (n-d_n)\ln\left(\frac{1-p_1}{1-p_0}\right) < \ln\left(\frac{1-\beta}{\alpha}\right)$$

After simplifications, we can obtain

$$s_b n + b_1 < d_n < s_b n + b_0 \tag{3.88}$$

where

$s_b = -\frac{\log\left(\frac{1-p_1}{1-p_0}\right)}{\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$, $b_1 = \frac{\log\left(\frac{\beta}{1-\alpha}\right)}{\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$, $b_0 = \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$. Thus the acceptance line and rejection line are, respectively, as follows:

$$\text{Acceptance line} = s_b n + b_1 \tag{3.89}$$

and

$$\text{Rejection line} = s_b n + b_0. \tag{3.90}$$

Similar tests can be constructed for other distribution parameters following the same general scheme.

Example 3.41

The following data were drawn one observation at a time in the order records:

g b g g b g b g b g g b g b g.

where b denotes a defective item and g denotes a good item. The experiment was performed to test the following hypothesis:

H_0: $p = p_0 = 0.10$ versus H_1: $p = p_1 = 0.2$

where p denotes the proportion defective items in the population. It is desired to reject H_0 when it is true with probability 0.05 and to accept H_0 when H_1 is true with probability 0.20. Using the sequential testing plan, we wish to determine whether we would accept or reject a lot on the basis of the observations above.

Given $p_0 = 0.10$, $p_1 = 0.20$, $\alpha = 0.05$, $\beta = 0.20$. From Eq. (3.88) we have

$$s_b = -\frac{\ln\left(\frac{1-p_1}{1-p_0}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)} = -\frac{\ln\left(\frac{1-0.20}{1-0.10}\right)}{\ln\left(\frac{0.2(1-0.10)}{0.10(1-0.20)}\right)} = 0.1452$$

$$b_1 = \frac{\ln\left(\frac{\beta}{1-\alpha}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)} = \frac{\ln\left(\frac{0.2}{1-0.05}\right)}{\ln\left(\frac{0.2(1-0.1)}{0.1(1-0.2)}\right)} = -1.9214,$$

$$b_0 = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\ln\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)} = \frac{\ln\left(\frac{1-0.2}{0.05}\right)}{\ln\left(\frac{0.2(1-0.1)}{0.1(1-0.2)}\right)} = 3.4190,$$

Thus,

$$s_b n + b_1 < d_n < s_b n + b_0$$

Table 3.13 The sampling procedure

n	x_n	d_n	Acceptance #	Rejection #
1	0	0	–	–
2	1	1	–	–
3	0	1	–	–
4	0	1	–	–
5	1	2	–	5
6	0	2	–	5
7	1	3	–	5
8	0	3	–	5
9	1	4	–	5
10	0	4	–	5
11	0	4	–	6
12	1	5	–	6
13	0	5	–	6
14	1	6	0	6
15	0	6	0	6

$$0.1452n - 1.9214 < d_n < 0.1452n + 3.4190$$

The acceptance line is: $0.1452n - 1.9214$.

The rejection line is: $0.1452n + 3.4190$.

From Table 3.13, the cumulative number of defective units $d_{14} = 6$ falls above the rejection line when $n = 14$. Therefore, the sampling is stopped at the 14th observation and that the lot would be rejected. That is, reject the hypothesis H_0: $p = 0.10$.

3.14 Bayesian Methods

The Bayesian approach to statistical inference is based on a theorem first presented by the Reverend Thomas Bayes. To demonstrate the approach, let X have a pdf $f(x)$, which is dependent on θ. In the traditional statistical inference approach, θ is an unknown parameter, and hence, is a constant. We now describe our prior belief in the value of θ by a pdf $h(\theta)$. This amounts to quantitatively assessing subjective judgment and should not be confused with the so-called objective probability assessment derived from the long-term frequency approach. Thus, θ will now essentially be treated as a random variable θ with pdf $h(\theta)$.

Consider a random sample $X_1, X_2, \ldots, X_n$ from $f(x)$ and define a statistic Y as a function of this random sample. Then there exists a conditional pdf $g(y \mid \theta)$ of Y for a given θ. The joint pdf for y and θ is

$$f(\theta, y) = h(\theta)g(y|\theta)$$

If θ is continuous, then

$$f_1(y) = \int_{\theta} h(\theta)g(y|\theta)d\theta$$

is the marginal pdf for the statistic y. Given the information y, the conditional pdf for θ is

$$\begin{aligned} k(\theta|y) &= \frac{h(\theta)g(y|\theta)}{f_1(y)} \quad \text{for } f_1(y) > 0 \\ &= \frac{h(\theta)g(y|\theta)}{\int_{\theta} h(\theta)g(y|\theta)d\theta} \end{aligned}$$

If θ is discrete, then

$$f_1(y) = \sum_k P(\theta_k)P(y|\theta_k)$$

and

$$P(\theta_i|y_i) = \frac{P(\theta_k)P(y_i|\theta_i)}{\sum_k P(\theta_k)P(y_j|\theta_k)}$$

where $P(\theta_j)$ is a prior probability of event θ_i and $P(\theta_j|\, y_j)$ is a posterior probability of event y_j given θ_i. This is simply a form of Bayes' theorem. Here, $h(\theta)$ is the prior pdf that expresses our belief in the value of θ before the data ($Y = y$) became available. Then $k(\theta \mid y)$ *is* the posterior pdf of given the data ($Y = y$).

Note that the change in the shape of the prior pdf $h(\theta)$ to the posterior pdf $k(\theta \mid y)$ due to the information is a result of the product of $g(y \mid \theta)$ and $h(\theta)$ because $f_1(y)$ is simply a normalization constant for a fixed y The idea in reliability is to take "prior" data and combine it with current data to gain a better estimate or confidence interval or test than would be possible with either singularly. As more current data is acquired, the prior data is "washed out" (Pham 2000).

Case 1: Binomial Confidence Limits—Uniform Prior. Results from ten missile tests are used to form a one-sided binomial confidence interval of the form

$$P[R \geq R_L] = 1 - \alpha$$

From subsection 3.11.3, we have

Table 3.14 Lower limits as a function of the number of missile test successes

K	RL	Exact level
10	0.79	0.905
9	0.66	0.904
8	0.55	0.900
7	0.45	0.898
6	0.35	0.905

$$\sum_{i=k}^{10} \binom{10}{i} R_L^i (1 - R_L)^{10-i} = \alpha$$

Choosing $\alpha = 0.1$, lower limits as a function of the number of missile test successes are shown in Table 3.14. Assume from previous experience that it is known that the true reliability of the missile is somewhere between 0.8 and 1.0 and furthermore that the distribution through this range is uniform. The prior density on R *is* then.

$g(R) = 5\ 0.8 < R < 1.0$

From the current tests, results are k successes out of ten missile tests, so for the event A that contained k successes:

$$P(A|R) = \binom{10}{k} R^k (1 - R)^{10-k}$$

Applying Bayes' theorem, we obtain

$$g(R|A) = \frac{g(R)P(A|R)}{\int_R g(R)P(A|R)dR}$$

$$= \frac{5\binom{10}{k} R^k (1 - R)^{10-k}}{\int_{0.8}^{1.0} 5\binom{10}{k} R^k (1 - R)^{10-k} dR}$$

For the case of $k = 10$,

$$g(R|A) = \frac{R^{10}}{\int_{0.8}^{1.0} R^{10} dR}$$

$$\frac{11\ R^{10}}{0.914} = 12.035\ R^{10}$$

To obtain confidence limits incorporating the "new" or current data,

Table 3.15 Comparison between limits applying the Bayesian method and those that do not

K	R_L (uniform [0.8,1] prior)	R_L (no prior)	Exact level
10	0.855	0.79	0.905
9	0.822	0.66	0.904
8	0.812	0.55	0.900
7	0.807	0.45	0.898
6	0.805	0.35	0.905

$$\int_{R_L}^{1.0} g(A|R)dR = 0.9$$
$$\int_{R_L}^{1.0} 12.035R^{10}dR = 0.9$$

After simplifications, we have

$$R_L^{11} = 0.177,\ R_L = 0.855.$$

Limits for the 10/10, 9/10, 8/10, 7/10, and 6/10 cases employing the Bayesian method are given in Table 3.15 along with a comparison with the previously calculated limits not employing the prior assumption. Note that the lower limit of 0.8 on the prior cannot be washed out.

Case 2: Binomial Confidence Limits—Beta Prior. The prior density of the beta function is

$$g(R) = \frac{(\alpha+\beta+1)!}{\alpha!\,\beta!}R^{\alpha}(1-R)^{\beta}$$

The conditional binomial density function is

$$P(A|R) = \binom{10}{i}R^{i}(1-R)^{10-i}$$

Then we have

$$g(R|A) = \frac{\frac{(\alpha+\beta+1)!}{\alpha!\,\beta!}R^{\alpha}(1-R)^{\beta}\binom{10}{k}R^{k}(1-R)^{10-k}}{\frac{(\alpha+\beta+1)!}{\alpha!\,\beta!}\int_0^1 R^{\alpha}(1-R)^{\beta}\binom{10}{k}R^{k}(1-R)^{10-k}dR}.$$

After simplifications, we obtain

$$g(R|A) = \frac{R^{\alpha+k}(1-R)^{\beta+10-k}}{\int_0^1 R^{\alpha+k}(1-R)^{\beta+10-k}dR}.$$

Multiplying and dividing by

$$\frac{(\alpha+\beta+11)!}{(\alpha+\beta)!(\beta+10-k)!}$$

puts the denominator in the form of a beta function with integration over the entire range, and hence, equal to 1. Thus,

$$g(R|A) = \begin{pmatrix} \alpha+\beta+10 \\ \alpha+k \end{pmatrix} R^{\alpha+k}(1-R)^{\beta+10-k}$$

which again is a beta density function with parameters

$$(\alpha+k) = \alpha' \text{ and } (\beta+10-k) = \beta'$$

Integration over $g(R|A)$ from R_L to 1.0 with an integral set to $1-\alpha$ and a solution of R_L will produce $100(1-\alpha)\%$ lower confidence bounds on R, that is,

$$\int_{R_L}^{1.0} g(R|A)dR = 1-\alpha.$$

Case 3: Exponential Confidence Limits—Gamma Prior. For this situation, assume interest is in an upper limit on the exponential parameter X. The desired statement is of the form

$$p[\lambda < \lambda_U] = 1-\alpha$$

If 1000 h of test time was accrued with one failure, a 90% upper confidence limit on λ would be

$$p[\lambda < 0.0039] = 0.9$$

From a study of prior data on the device, assume that λ has a gamma prior density of the form

$$g(\lambda) = \frac{\lambda^{n-1}\, e^{-\frac{\lambda}{\beta}}}{(n-1)!\,\beta^n}.$$

With an exponential failure time assumption, the current data in terms of hours of test and failures can be expressed as a Poisson, thus,

$$p(A|\lambda) = \frac{(\lambda T)^n \, e^{-\lambda T}}{r!}$$

where $n =$ number of failures, $T =$ test time, and $A =$ event which is r failures in T hours of test. Applying Bayes' results, we have

$$g(\lambda|A) = \frac{\frac{\lambda^{n-1} \, e^{-\frac{\lambda}{\beta}}}{(n-1)! \, \beta^n} \frac{(\lambda T)^r e^{-\lambda T}}{r!}}{\int_{\lambda=0}^{\infty} \frac{\lambda^{n-1} \, e^{-\frac{\lambda}{\beta}}}{(n-1)! \, \beta^n} \frac{(\lambda T)^r e^{-\lambda T}}{r!} d\lambda}$$
$$= \frac{\lambda^{n+r-1} e^{-\lambda\left(\frac{1}{\beta}+T\right)}}{\int_0^{\infty} \lambda^{n+r-1} \, e^{-\lambda\left(\frac{1}{\beta}+T\right)} d\lambda}.$$

Note that

$$\int_0^{\infty} \lambda^{n+r-1} e^{-\lambda\left(\frac{1}{\beta}+T\right)} d\lambda = \frac{(n+r-1)!}{\left(\frac{1}{\beta}+T\right)^{n+r}}$$

Hence,

$$g(\lambda|A) = \frac{\lambda^{n+r-1} e^{-\lambda\left(\frac{1}{\beta}+T\right)} \left(\frac{1}{\beta}+T\right)^{n+r}}{(n+r-1)!}$$

Thus, $g(\lambda|A)$ is also a gamma density with parameters $(n+r-1)$ and $\frac{1}{\left(\frac{1}{\beta}+T\right)}$. This density can be transformed to the χ^2 density with $2(n+r)$ degree of freedom by the following change of variable. Let

$$\lambda' = 2\lambda \left(\frac{1}{\beta}+T\right)$$

then

$$d\lambda = \frac{1}{2}\left(\frac{1}{\frac{1}{\beta}+T}\right) d\lambda'$$

We have

$$h(\lambda'|A) = \frac{(\lambda')^{\frac{2(n+r)}{2}-1} e^{-\frac{\lambda'}{2}}}{\left[\frac{2(n+r)}{2}-1\right]! \; 2^{\frac{2(n+r)}{2}}}$$

To obtain a 100(1−α)% upper confidence limit on λ, solve for λ' in the integral

$$\int_0^{\lambda'} h(s|A)ds = 1 - \alpha$$

and convert λ' to λ via the above transformation.

Example 3.42
Given a gamma prior with $n = 2$ and $\beta = 0.0001$ and current data as before (i.e., 1000 h of test with one failure), the posterior density becomes.

$$g(\lambda|A) = \frac{\lambda^2\, e^{-\lambda(11,000)}(11,000)^2}{2}$$

converting to χ^2 via the transformation $\lambda' = 2\lambda$ (11,000) and

$$h(\lambda'|A) = \frac{(\lambda')^{\frac{6}{2}-1}e^{-\frac{\lambda'}{2}}}{\left[\frac{6}{2}-1\right]!\, 2^{\frac{6}{2}}}$$

which is χ^2 with six degrees of freedom. Choosing $\alpha = 0.1$ then

$$\chi^2_{6,1-\alpha} = \chi^2_{6,0.9} = 10.6$$

and

$$p[\lambda' < 10.6] = 0.9$$

But $\lambda' = 2\lambda(11{,}000)$, hence,

$$p\left[\lambda' = 2\lambda(11,000) < 10.6\right] = 0.9$$

or

$$p[\lambda < 0.0005] = 0.9$$

The latter limit conforms to 0.0039 derived without the use of a prior density, i.e., an approximate eight fold improvement. The examples above involved the development of tighter confidence limits where a prior density of the parameter could be utilized.

In general, for legitimate applications and where prior data are available, employment of Bayesian methods can reduce cost or give results with less risk for the same dollar value (Pham 2000).

3.15 Statistical Model Selection

In this section we discuss some common criteria that can be used for distribution and model selections (Pham 2014, 2019,2020). Let y_i be the observed data value and $\hat{y}_i$ is the fitted value from the fit for $i = 1,2,\ldots,n$; and n and k are the number of observations and number of estimated parameters, respectively.

Define the sum of squared error (SSE) and total sum of squares (SST) as follows:

$$SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{3.91}$$

and

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

where

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

***Mean Squared Error* (*MSE*).** The mean squared error (MSE) measures the total deviation of the response values from the fitted response values and is defined as:

$$\text{MSE} = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n-k} = \frac{SSE}{n-k} \tag{3.92}$$

where y_i is the observed data value; $\hat{y}_i$ is the fitted value from the fit for $i = 1,2,\ldots,n$; and n and k are the number of observations and number of estimated parameters, respectively. SSE is the sum of squared error. Note that MSE considers the penalty term with respect to the degrees of freedom when there are many parameters and consequently assigns a larger penalty to a model with more parameters. The smaller the value MSE value, the better the model fit.

***Root Mean Squared Error* (*RMSE*).** The RMSE is the square root of the variance of the residuals or the MSE and is given as

$$RMSE = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k}} \tag{3.93}$$

The RMSE indicates the absolute fit of the model to the data–how close the observed data points are to the model's predicted values. Lower values of RMSE indicate better fit.

***Coefficient of Determinations* (R^2).** R^2 is the square of the correlation between the response values and the predicted response values. It measures the amount of variation accounted for the fitted model. The R^2 is defined as

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} \tag{3.94}$$

R^2 assumes every independent variable in the model explains the variation in the dependent variable. It gives the percentage of explained variation as if all independent variables in the model affect the dependent variable. It is always increase with model size. The larger R^2, the better is the model's performance.

***Adjusted R^2*.** Adjusted R^2 takes into account the number of estimated parameters in the model and is defined as:

$$R^2_{adj} = 1 - \left(\frac{n-1}{n-k}\right)\left(1 - R^2\right) \tag{3.95}$$

where n and k are the number of observations and number of estimated parameters, respectively. The adjusted R^2 gives the percentage of variation explained by only those independent variables that actually affect the dependent variable. The larger adjusted R^2, the better is the model's goodness-of-fit.

***The predictive-ratio risk* (*PRR*).** PRR measures the total deviation of the response values from the fit to the response values against the fitted values, and is defined as (Pham 2006; Pham and Deng 2003):

$$PRR = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)^2 \tag{3.96}$$

***The predictive-power* (*PP*).** PP measures the total deviation of the response values from the fit to the response values against the response values, and is defined as follows:

$$PP = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{y_i}\right)^2 \tag{3.97}$$

For all these three criteria—MSE, PRR, and PP—the smaller the value, the better the model fits.

***Normalized-Rank Euclidean Distance criterion* (*RED*).** In general, let s denotes the number of models with d criteria, C_{ij1} represents the ranking based on specified criterion of model i with respect to (w.r.t.) criteria j, and C_{ij2} the criteria value of model i w.r.t. criteria j where $i = 1, 2,\ldots,s$ and $j = 1, 2,\ldots,d$. The normalized-rank Euclidean distance value, D_i, for $i = 1, 2,\ldots, s$, measures the distance of the two-dimensional normalized criteria from the origin for *ith* model using Euclidean distance function (Pham 2019). The RED criteria function is defined as follows:

$$D_i = \sum_{j=1}^{d} \left\{ \left(\sqrt{\left[\sum_{k=1}^{2} \left(\frac{C_{ijk}}{\sum_{i=1}^{s} C_{ijk}} \right)^2 \right]} \right) w_j \right\} \tag{3.98}$$

where s = total number of models.

d = total number of criteria.

w_j = the weight of the jth criteria for $j = 1,2,\ldots,d$

$$k = \begin{cases} 1 \text{ represent criteria j value} \\ 2 \text{ represent criteria j ranking.} \end{cases}$$

Thus, the smaller the RED value, D_i, it represents the better rank as compare to higher RED value.

***Akaike Information Criterion* (*AIC*).** The AIC is one of the most common criteria has been used in past years to choose the best model from a set of candidate models. It is defined as (Akaike 1973):

$$AIC = -2\log(L) + 2k \tag{3.99}$$

where L is the maximum value of the likelihood function for the model and k is the number of estimated parameters in the model. To select the best model, practitioners should choose the model that minimizes AIC. In other words, the lower value of AIC indicates better goodness-of-fit. By adding more parameters in the model, it improves the goodness of the fit but also increases the penalty imposed by adding more parameters.

***Bayesian Information Criterion* (*BIC*).** The BIC was introduced by Schwarz (1978) also known as Schwarz criterion, and is defined as

$$BIC = -2\log(L) + k\log(n) \tag{3.100}$$

As can be seen from Eqs. (3.5) and (3.6), the difference between these two criteria, BIC and AIC, is only in the second term which depends on the sample size n that show how strongly they impacts the penalty of the number of parameters in the model. With AIC the penalty is 2 k whereas with BIC the penalty is k.log(n) that penalizes large models. As n increases, the BIC tends to favor simpler models than the AIC as k smaller. This implies when n > 8, k·log(n) exceeds 2 k.

***Second Order Information Criterion* (*AICc*).** When the sample size is small, there is likely that AIC will select models that include many parameters. The second order information criterion, often called AICc, takes into account sample size by increasing the relative penalty for model complexity with small data sets. It is defined as:

$$AIC_c = -2\log(L) + 2k + \frac{2k(k+1)}{n-k-1} \tag{3.101}$$

where n denotes the sample size and k denotes the number of estimated parameters. As n gets larger, AICc converges to AIC so there's really no harm in always using AICc regardless of sample size.

***Pham's information criterion* (*PIC*).** In general, the adjusted R^2 attaches a small penalty for adding more variables in the model. The difference between the adjusted R^2 and R^2 is usually slightly small unless there are too many unknown coefficients in the model to be estimated from too small a sample in the presence of too much noise. Pham (2019) presents a criterion, called Pham's information criterion (PIC), by taking into account a larger the penalty when adding too many coefficients in the model when there is too small a sample, is as follows:

$$\text{PIC} = \text{SSE} + k\left(\frac{n-1}{n-k}\right) \tag{3.102}$$

where n is the number of observations in the model, k is the number of estimated parameters or $(k-1)$ explanatory variables in the model, and SSE is the sum of squared error as given in Eq. (3.91).

***Pham's criterion* (*PC*).** Pham (2020) recently introduces a criterion, called Pham's criterion (PC), that measures the tradeoff between the uncertainty in the model and the number of parameters in the model by slightly increasing the penalty when each time adding parameters in the model when there is too small a sample. The criteria is as follows (Pham 2020a):

$$\begin{aligned} PC &= \left(\frac{n-k}{2}\right)\log\left(\frac{SSE}{n}\right) + k\left(\frac{n-1}{n-k}\right) \\ \text{where } SSE &= \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \end{aligned} \tag{3.103}$$

Table 3.16 presents a summary of criteria for model selection.

3.15.1 Applications

In this section we demonstrate a couple of real applications such as advertising budget products and heart blood pressure health using multiple regression models to illustrate the model selection.

Application 1: Advertising Budget Analysis.

In this application, we use the advertising budget data set [Advertising] to illustrate the model selection criteria where the sales for a particular product is a dependent variable of multiple regression and the three different media channels such as TV, Radio, and Newspaper are independent variables as shown in Table 3.17. The advertising dataset consists of the sales of a product in 200 different markets (200 rows),

Table. 3.16 A summary of some criteria model selection (Pham 2020a)

No	Criteria	Formula	Brief description
1	SSE	$\text{SSE} = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$	Measures the total deviations between the estimated values and the actual data
2	MSE	$\text{MSE} = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n-k}$	Measures the difference between the estimated values and the actual data
3	RMSE	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{n-k}}$	The square root of the MSE
4	R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$	Measures the amount of variation accounted for the fitted model
5	Adj R^2	$R^2_{adj} = 1 - \left(\frac{n-1}{n-k}\right)\left(1 - R^2\right)$	Take into account a small penalty for adding more variables in the model
6	AIC	$AIC = -2\log(L) + 2k$	Measure the goodness of the fit considering the penalty of adding more parameters
7	BIC	$BIC = -2\log(L) + k\log(n)$	Same as AIC but the penalty term will also depend on the sample size
8	AICc	$AIC_c = -2\log(L) + 2k + \frac{2k(k+1)}{n-k-1}$	AICc takes into account sample size by increasing the relative penalty for model complexity with small data sets
9	PIC	$\text{PIC} = SSE + k\left(\frac{n-1}{n-k}\right)$ where $SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$	Take into account a larger the penalty when there is too small a sample but too many parameters in the model
10	PRR	$PRR = \sum_{i=1}^{n} \left(\frac{\hat{m}(t_i) - y_i}{\hat{m}(t_i)}\right)^2$	Measures the distance of model estimates from the actual data against the model estimate
11	PP	$PP = \sum_{i=1}^{n} \left(\frac{\hat{m}(t_i) - y_i}{y_i}\right)^2$	Measures the distance of model estimates from the actual data against the actual data

(continued)

Table. 3.16 (continued)

No	Criteria	Formula	Brief description
12	PC	$PC = \left(\frac{n-k}{2}\right)\log\left(\frac{SSE}{n}\right) + k\left(\frac{n-1}{n-k}\right)$ where $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Increase slightly the penalty each time adding parameters in the model when there is too small a sample

Table 3.17 Advertising Budget Data in 200 different markets

TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75	7.2
57.5	32.8	23.5	11.8
120.2	19.6	11.6	13.2
8.6	2.1	1	4.8

together with advertising budgets for the product in each of those markets for three different media channels: TV, Radio and Newspaper. The sales are in thousands of units and the budget is in thousands of dollars. Table 3.17 shows just the first few rows of the advertising budget data set.

Pham (2019) discusses the results of the linear regression model using this advertising data. Figures 3.6 and 3.7 present the data plot and the correction coefficients between the pairs of variables of the advertising budget data, respectively. It shows that the pair of Sales and TV advertising have the highest positive correlation. From Table 3.18, the values of R^2 with all three variables and just two variables (TV and Radio advertisings) in the model are the same. This implies that we can select the model with two variables (TV and Radio) in the regression (Pham 2019). Based on the PIC criterion, the model with the two advertising media channels (TV and Radio) is the best model from a set of seven candidate models as shown in Table 3.18.

Application 2: Heart Blood Pressure Health Analysis.

The heart blood pressure health data (Pham 2019) consists of the heart rate (pulse), systolic blood pressure and diastolic blood pressure in 86 days with 2 data points measured each day (172 rows) is used in this application. Blood pressure (BP) is one of the main risk factors for cardiovascular diseases. BP is the force of blood pushing against your artery walls as it goes through your body [Blood pressure]. The Systolic BP is the pressure when the heart beats—while the heart muscle is contracting (squeezing) and pumping oxygen-rich blood into the blood vessels. Diastolic BP is

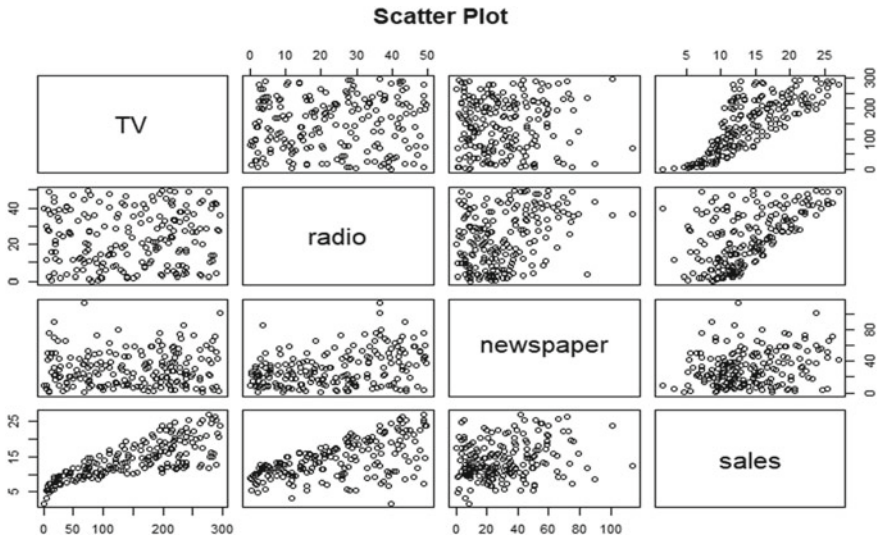

Fig. 3.6 The scatter plot of the advertising data with four variables

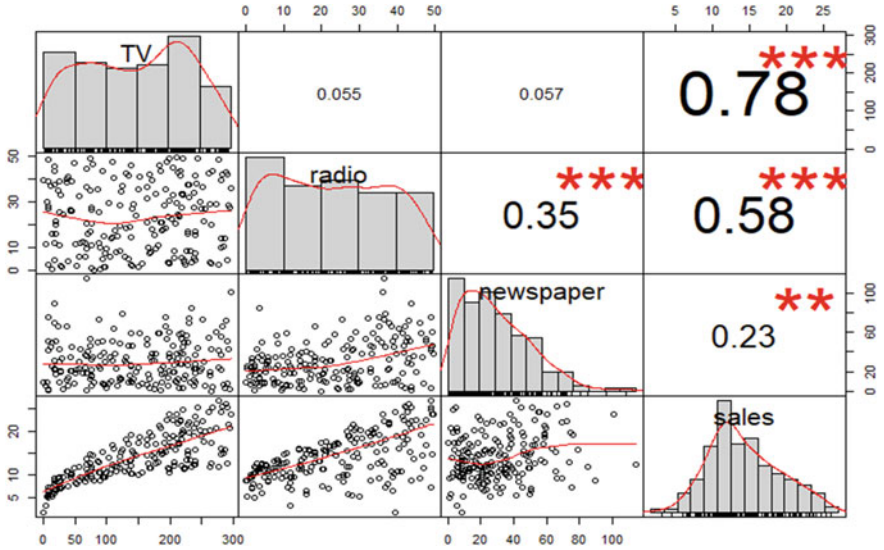

Fig. 3.7 The Correlation Coefficient between the Variables

the pressure on the blood vessels when the heart muscle relaxes. The diastolic pressure is always lower than the systolic pressure [Systolic]. The Pulse or Heart rate measures the heart rate by counting the number of beats per minute (BPM). The first few rows of the data set are shown in Table 3.19. In the table, the first row of

Table 3.18 Criteria values of independent variables (TV, Radio, Newspaper) of regression models (X_1, X_2, and X_3 be denoted as the TV, radio and newspaper, respectively)

Criteria	X_1, X_2, X_3	X_1, X_2	X_1, X_3	X_2, X_3	X_1	X_2	X_3
MSE	2.8409	2.8270	9.7389	18.349	10.619	18.275	25.933
AIC	782.36	780.39	1027.8	1154.5	1044.1	1152.7	1222.7
AICc	782.49	780.46	1027.84	1154.53	1044.15	1152.74	1222.73
BIC	795.55	790.29	1037.7	1164.4	1050.7	1159.3	1229.3
RMSE	1.6855	1.6814	3.1207	4.2836	3.2587	4.2750	5.0925
R^2	0.8972	0.8972	0.6458	0.3327	0.6119	0.3320	0.0521
Adjusted R^2	0.8956	0.8962	0.6422	0.3259	0.6099	0.3287	0.0473
PIC	5.7467	**4.7118**	6.1512	7.3141	5.2688	6.2850	7.1026

Table 3.19 Sample heart blood pressure health data set of an individual in 86 day interval (Pham 2019)

Day	Time	Systolic	Diastolic	Pulse
5	0	154	99	71
	1	144	94	75
6	0	139	93	73
6	1	128	76	85
7	0	129	73	78
7	1	125	65	74
1	0	129	80	70
1	1	130	83	72
2	0	144	83	74
2	1	124	87	84
3	0	120	77	73
3	1	124	70	80

the data set can be read as follows: on a Thursday morning, the high blood, low blood and heart rate measurements were 154, 99, and 71, respectively. Similarly, on a Thursday afternoon (i.e., the second row of the data set in Table 3.19), the high blood, low blood and heart rate measurements were 144, 94, and 75, respectively.From Fig. 3.8, the systolic BP and diastolic BP have the highest correlation. The model with only Systolic blood pressure variable seems to be the best model from the set of seven candidate models based on PIC and BIC criteria as shown in Table 3.20.

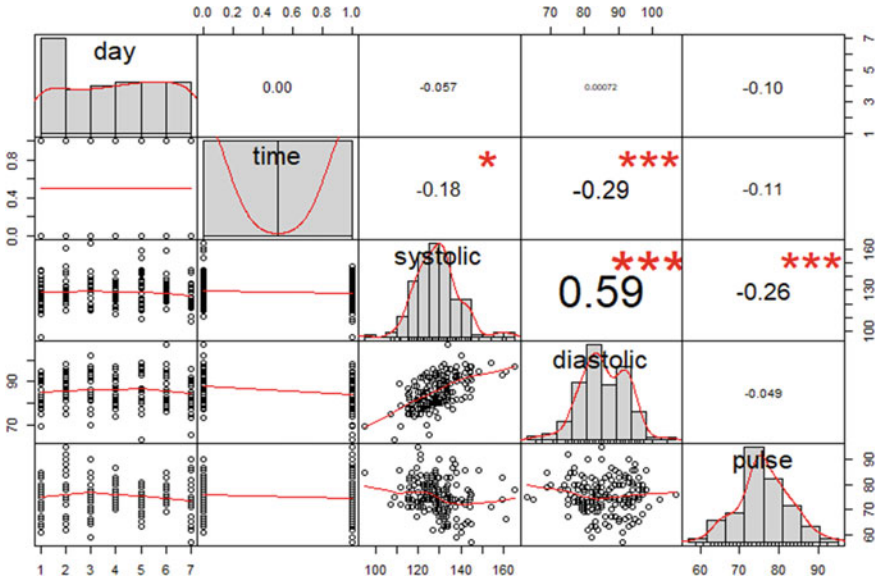

Fig. 3.8 The correlation coefficient between the variables

Table 3.20 Criteria values of variables (day, systolic, diastolic) of regression models (X_1, X_2, and X_3 be denoted as the day, systolic, diastolic, respectively)

Criteria	X_1, X_2, X_3	X_1, X_2	X_1, X_3	X_2, X_3	X_1	X_2	X_3
MSE	43.1175	43.7381	47.0101	43.5859	46.8450	44.1352	47.2311
AIC	1141.463	1142.942	1155.351	1142.342	1153.76	1143.511	1155.172
BIC	1154.053	1152.384	1164.793	1151.784	1160.055	1149.806	1161.467
RMSE	6.5664	6.6135	6.8564	6.6020	6.8443	6.6434	6.8725
R^2	0.09997	0.0816	0.01287	0.08477	0.0105	0.0678	0.00236
Adj R^2	0.08389	0.0707	0.00119	0.07394	0.00469	0.0623	−0.00351
PIC	10.6378	9.6490	9.8919	9.6375	8.8561	**8.6552**	8.8843

3.16 Problems

1. Let $X_1, X_2, \ldots, X_n$ represent a random sample from the Poisson distribution having pdf

   $f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ for x = 0, 1, 2,… and $\lambda \geq 0$.
   Find the maximum likelihood estimator $\hat{\lambda}$ of λ.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with a discrete pdf

$$P(x) = p^x(1-p)^{1-x} \quad x = 0, 1 \text{ and } 0 < p < 1$$

where p is the parameter to be estimated. Find the maximum likelihood estimator $\hat{p}$ of p.

3. Assume that $X_1, X_2, \ldots, X_n$ represent a random sample from the Pareto distribution, that is,

$$F(x; \lambda, \theta) = 1 - \left(\frac{\lambda}{x}\right)^{\theta} \quad \text{for } x \geq \lambda,\ \lambda > 0,\ \theta > 0$$

This distribution is commonly used as a model to study incomes. Find the maximum likelihood estimators of λ and θ.

4. Let $Y_1 < Y_2 < \ldots < Y_n$ be the order statistics of a random sample $X_1, X_2, \ldots, X_n$ from the distribution with pdf

$$f(x; \theta) = 1 \quad \text{if } \theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2},\ -\infty < \theta < \infty$$

Show that any statistic $h(X_1, X_2, \ldots, X_n)$ such that

$$Y_n - \frac{1}{2} \leq h(X_1, X_2, ..., X_n) \leq Y_1 + \frac{1}{2}$$

is a maximum likelihood estimator of θ. What can you say about the following functions?

(a) $\dfrac{(4Y_1 + 2Y_n + 1)}{6}$

(b) $\dfrac{(Y_1 + Y_n)}{2}$

(c) $\dfrac{(2Y_1 + 4Y_n - 1)}{6}$

5. The lifetime of transistors is assumed to have an exponential distribution with pdf

$$f(t; \theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \quad \text{for t} \geq 0,\ \theta > 0$$

A random sample of size n *is* observed. Determine the following:

(a) The maximum likelihood estimator of θ.
(b) The MLE of the transistor reliability function, $\hat{R}(t)$, of

$$R(t) = e^{-\frac{t}{\theta}}$$

6. Suppose that $X_1, X_2,\ldots, X_n$ are independent random variable, each with the uniform distribution on $[c - d, c + d]$ where c is unknown and d is known $(-\infty < c < \infty,\ d > 0)$. Find the maximum likelihood estimator of c.
7. Suppose that $X_1, X_2,\ldots, X_n$ are independent random variable, each with the uniform distribution on $[c - d, c + d]$ where c and d are both unknown $(-\infty < c < \infty,\ d > 0)$.

   (a) Find the maximum likelihood estimators of c and d.
   (b) Given the following failure time data: 20, 23, 25, 26, 28, 29, 31, 33, 34, and 35 days. Assuming that the data follow a uniform distribution on $[c - d, c + d]$, use the maximum likelihood method to obtain the unknown parameters, c and d, of the uniform distribution.

8. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variable, each with the uniform distribution on $[-d, d]$ where d is positive and unknown. Find the maximum likelihood estimator of d.
9. Suppose on five consecutive days in a given month the number of customers who enter services at a printing shop were 48, 60, 78, 56, and 73 Test the null hypothesis that the expected numbers of customers per day were the same on the five days at the 5% level of significance using the Chi-square test.
10. Suppose that a die is rolled 120 times and the number of times each face comes up is recorded. The following results are obtained:

Face	1	2	3	4	5	6
n_i	15	21	20	15	26	23

Test whether the die is fair at the 5% level of significance using the Chi-square test.

11. Suppose that a die is rolled 250 times and the number of times each face comes up is recorded. The following results are obtained:

Face	1	2	3	4	5	6
n_i	45	37	60	55	29	24

Test whether the die is fair at the 5% level of significance using the Chi-square test.

12. The proportion of components produced by a manufacturing process from last year is as follows: 3% scrapped, 6% reworked, and 91% acceptable. This year, inspection of 500 units showed that 20 units must be scrapped and 25 units can be reworked. Can we say that the results this year are consistent with the last year data at the 5% level of significance using the Chi-square test.
13. An engineer obtained the following data that shows the cycle time in hours for the assembly of a certain electronic product:

Cycle time (hours)	Frequency
2.45	4
2.55	6
2.58	15
2.64	8
2.73	2
2.80	29
2.85	7
2.92	13
2.98	19
3.20	12

The engineer concludes that these data might represent like a sample from a normal population at the 5% level of significance. What is your opinion. Is the engineer correct? (Hint: using the KS test).

14. 27 units were placed on life test and the test was run until all units failed. The observed failure times $t_1, t_2, \ldots, t_{27}$ were give below:

4.6	12.7	20.5
5.4	13.2	20.9
5.8	13.5	21.5
6.7	13.9	22.7
7.3	14.7	23.6
7.9	17.5	24.9
8.7	17.6	25.4
9.9	19.3	25.9
12.5	20.3	35.3

Test the hypothesis that the underlying distribution of life is exponential at the 5% level of significance using the Chi-square test.

15. Show that

$$I(\theta) \equiv E\left\{\left[\frac{\partial \ln f(x;\theta)}{\partial \theta}\right]^2\right\} = -E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right)$$

16. Suppose that we have k disjoint events $A_1, A_2, \ldots, A_k$ such that the probability of A_i is p_i for $i = 1,2,\ldots, k$ and $\sum_{i=1}^{k} p_i = 1$. Let's assume that among n independent trials there are $X_1, X_2, \ldots, X_k$ outcomes associated with $A_1, A_2, \ldots, A_k$, respectively. The joint probability that $X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k$ is given as follows by the likelihood function

$$L(x, p) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_1^{x_1} \ldots p_k^{x_k}$$

where $\sum_{i=1}^{k} x_i = n$. This is infact known as the multinomial distribution. Find the MLE of $p_1, p_2, \ldots, p_k$ by maximizing the likelihood function above.

17. Consider the binomial distribution with unknown parameter p given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 1, 2, \ldots, n;\ 0 \le p \le 1$$

where n = number of trials; k = number of successes; p = single trial probability of success. Find the maximum likelihood estimator of p.

18. (a) Find the maximum likelihood estimator of θ if $t_1, t_2, \ldots, t_n$ are independent observations from a population with the following probability density function

$$f(t; \theta) = \frac{\theta}{t^{\theta+1}} \quad \text{for } t \ge 1,\ \theta > 0$$

(b) Given the following failure time data: 200, 225, 228, 245, 250, 286, 290 h. Assuming that the data follow the above pdf, obtain the MLE of θ.

19. (a) Find the maximum likelihood estimator of λ if the failure times $t_1, t_2, \ldots, t_n$ are independent observations from a population with the following probability density function

$$f(t) = (1 + \lambda)t^{\lambda} \quad \text{for } 0 < t < 1.$$

(b) The failure times are: 0.2, 0.3, 0.35, 0.45, 0.5, 0.6, 0.7, 0.75, 0.8, and 0.95 h. Obtain the MLE of λ.

20. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables, each with the following probability density function:

$$f(x, \alpha, \beta) = \begin{cases} 0 & \text{if } x < \alpha \\ \frac{1}{\beta} e^{-\frac{(x-\alpha)}{\beta}} & \text{if } x \ge \alpha \end{cases}$$

where $-\infty < \alpha < \infty$ and $0 < \beta < \infty$ are both unknown.

(a) Find the maximum likelihood estimators (MLE) of α and β, say $\hat{\alpha}$ and $\hat{\beta}$, respectively.
(b) Find $E(\hat{\alpha})$ and $E(\hat{\beta})$.
(c) If $n = 7$ and $x_1 = 5.3$, $x_2 = 3.2$, $x_3 = 2.4$, $x_4 = 3.8$, $x_5 = 4.2$, $x_6 = 3.4$, $x_7 = 2.9$, find the MLE of α and β

21. Suppose on five consecutive days in a given month the number of customers who enter services at a printing shop were 48, 60, 78, 56, and 73
Test the null hypothesis that the expected numbers of customers per day were the same on the five days at the 5% level of significance using the Chi-square test.
22. Suppose that X is a discrete random variable with the following probability mass function:

X	P(X)
0	$\frac{2\theta}{3}$
1	$\frac{\theta}{3}$
2	$\frac{2(1-\theta)}{3}$
3	$\frac{(1-\theta)}{3}$

where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution:
3, 0, 2, 1, 3, 2, 1, 0, 2, 1.
What is the maximum likelihood estimate of θ?

23. Let X denote the proportion of allotted time that a randomly selected engineer spends working on a certain project, and suppose that the probability density function of X is

$$f(x) = \begin{cases} (\theta+1)x^{\theta} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

where $\theta > -1$.
A random sample of 10 engineers yielded data

$$x_1 = 0.92,\ x_2 = 0.79,\ x_3 = 0.90,\ x_4 = 0.65,\ x_5 = 0.86$$
$$x_6 = 0.47,\ x_7 = 0.73,\ x_8 = 0.97,\ x_9 = 0.94,\ x_{10} = 0.77$$

(a) Use the method of moments to obtain an estimator of θ, and then compute the estimate for this data.
(b) Obtain the maximum likelihood estimator of θ, and then compute the estimate for the given data.
24. Let X be uniformly distributed on the interval $[0, \theta]$ and the probability density function of X is

$$f(x) = \begin{cases} \dfrac{1}{\theta} & \text{for } 0 \leq x \leq \theta \\ 0 & \text{otherwise.} \end{cases}$$

where $\theta > 0$.

A random sample of 5 inspectors yielded data

$x_1 = 10.52$, $x_2 = 12.37$, $x_3 = 8.90$, $x_4 = 14.35$, $x_5 = 9.66$

Obtain an estimator of θ using (a) the method of moments and (b) the method of maximum likelihood, and then compute the estimate for this data.

25. Suppose that my waiting time for a train is uniformly distributed on the interval $[0, \theta]$, and that the results $x_1, x_2, \ldots, x_n$ of a random sample from this distribution have been observed. If my waiting times are 4.5, 5.3, 1.2, 7.4, 3.6, 4.8, and 2.9,

    (a) calculate the estimate of θ using the method of moments.
    (b) calculate the estimate of θ using the method of maximum likelihood.

26. (a) Find the method of moments estimate for θ based on a random sample of size n taken from the following probability density function

$$f(x) = (\theta + 1)x^{\theta} \quad \text{for} \quad 0 < x < 1.$$

    (b) Calculate the estimate of θ for the sample $x_1 = 0.7$, $x_2 = 0.4$, $x_3 = 0.8$, $x_4 = 0.5$

27. At time $x = 0$, twenty identical units are put on test. Suppose that the lifetime probability density function of each unit with parameter λ is given by

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0, \lambda > 0$$

The quality manager then leaves the test facility unmonitored. On his return 24 h later, the manager immediately terminates the test after noticing that $y = 15$ of the 20 units are still in operation (so five have failed). Determine the maximum likelihood estimate of λ.

28. At time $x = 0$, forty identical units are put on test. Suppose that the lifetime probability density function of each unit with parameter θ is given by

$$f(x) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} \text{ for } x \geq 0, \quad \theta > 0$$

The quality manager then leaves the test facility unmonitored. On his return 8 h later, the manager immediately terminates the test after noticing that 30 of the 40 units are still in operation (so ten have failed). Determine the maximum likelihood estimate of θ.

29. The following data were drawn one observation at a time in the order records:

g g b g g b g g g b g b g g b b b g g g.

where *b* denotes a defective item and *g* denotes a good item. The experiment was performed to test the following hypothesis:

H_0: $p = p_0 = 0.10$ versus H_1: $p = p_1 = 0.20$.

where p denotes the proportion defective items in the population. It is desired to reject H_0 when it is true with probability 0.05 and to accept H_0 when H_1 is true with probability 0.20. Using sequential testing plan, determine whether you would accept or reject a lot on the basis of the observations above.

30. The following data were drawn one observation at a time in the order records:

b g b b g b g b.

where b denotes a defective item and g denotes a good item. The experiment was performed to test the following hypothesis:

H_0: $p = p_0 = 0.10$ versus H_1: $p = p_1 = 0.25$.

where p denotes the proportion defective items in the population. It is desired to reject H_0 when it is true with probability 0.05 and to accept H_0 when H_1 is true with probability 0.10. Using sequential testing plan, determine whether you would accept or reject a lot on the basis of the observations above.

31. Let X represent a random variable of service time at a certain facility, and suppose that the cumulative distribution function of X is

$$F(x) = \begin{cases} 1 - e^{-\lambda x} - \lambda x e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

where $\lambda > 0$. A random sample of 10 customers yielded data.

$$x_1 = 12, \quad x_2 = 18, \quad x_3 = 8, \quad x_4 = 15, \quad x_5 = 11,$$
$$x_6 = 9, \quad x_7 = 8, \quad x_8 = 10, \quad x_9 = 18, x_{10} = 21$$

(a) Use the method of moments to obtain an estimator of λ, and then compute the estimate for this data.

(b) Obtain the maximum likelihood estimator of λ, and then compute the estimate for the given data.

32. Let X represent a random variable of service time at a certain facility, and suppose that the probability density function of X is

$$f(x) = 0.5\, \lambda^3 x^2 e^{-\lambda x} \qquad \text{for } x > 0, \lambda > 0.$$

A random sample of 5 customers yielded data
$x_1 = 20, x_2 = 16, x_3 = 15, x_4 = 18, x_5 = 25$
Use the method of moments to obtain an estimator of λ, and then compute the estimate for this data.

33. A study of the relationship between facility conditions at gasoline stations and aggressiveness in the pricing of gasoline reported the accompanying data below based on a sample of n = 441 stations. At the 0.10 significance level,

does the data suggest that facility conditions and pricing policy are independent of one another?

Observed Pricing Policy

Facility conditions	Aggressive	Neutral	Nonaggressive	Total
Substandard	24	15	17	56
Standard	52	73	80	205
Modern	58	86	36	180
Total	134	174	133	441

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Caski F (eds) The second international symposium proceedings on information theory. AkademiaiKiado, Budapest, Hungary, pp 267–281

Asher H, Feingold H (1984) Repairable systems reliability. Marcel Dekker, New York

Feller W (1966) An introduction to probability theory and its applications, vol II. Wiley, New York

Feller W (1968) An introduction to probability theory and its applications, vol 1, 3rd edn. Wiley, New York

Jelinski Z, Moranda PB (1972) Software reliability research. In: Freiberger W (ed) Statistical computer performance evaluation. Academic Press, New York

Musa JD, Iannino A, Okumoto K (1987) Software reliability: measurement, prediction, application. McGraw-Hill, New York

Nelson SC, Haire MJ, Schryver JC (1992) Network-simulation modeling of interactions between maintenance and process systems. In: Proceedings of the 1992 annual reliability and maintainability symposium, Las Vegas, NV, USA,1992 Jan 21–23, pp 150–156

Pham H (2000) Software reliability. Springer, Singapore

Pham H (2002) A Vtub-shaped hazard rate function with applications to system safety. Int J Reliab Appl 3(1):1–16

Pham H (2006) System software reliability. Springer

Pham H (2020) On estimating the number of deaths related to Covid-19. Mathematics 8

Pham H (2014) A new software reliability model with Vtub-shaped fault-detection rate and the uncertainty of operating environments. Optimization 63(10):1481–1490

Pham H (2019) A new criteria for model selection. Mathematics 7(12):2019

Pham H, Deng C (2003) Predictive-ratio risk criterion for selecting software reliability models. In: Proceedings of the ninth international conference on reliability and quality in design

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 1978(6):461–464

Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions. Ann Math Stat 19:279–281

Zhao M (2003) Statistical reliability change-point estimation models. In: Pham H (ed) Handbook of reliability engineering. Springer, pp 157–163.