

Chapter 9

Estimation and Modelling

Abstract Earlier in Chaps. 3 and 7, several types of models for lifetime data were discussed through their quantile functions. These will be candidate distributions in specific situations. The selection of one of them or a new one is dictated by how well it can justify the data generating mechanisms and satisfy well other criteria like goodness of fit. Once the question of an initial choice of the model is resolved, the problem is then to test its adequacy against the observed data. This is accomplished by first estimating the parameters of the model and then carrying out a goodness-of-fit test. This chapter addresses the problem of estimation as well as some other modelling aspects.

In choosing the estimates, our basic objective is to get estimated values that are as close as possible to the true values of the model parameters. One method is to seek estimate that match the basic characteristics of the model with those in the sample. This includes the method of percentiles and the method of moments that involve the conventional moments, L -moments and probability weighted moments. These methods of estimation are explained along with a discussion of the properties of these estimates. In the quantile form of analysis, the method of maximum likelihood can also be employed. The approach of this method, when there is no tractable distribution function, is described. Many functions required in reliability analysis are estimated by nonparametric methods. These include the quantile function itself and other functions such as quantile density function, hazard quantile function and percentile residual quantile function. We review some important results in these cases that furnish the asymptotic distribution of the estimates and the proximity of the proposed estimates to the true values.

9.1 Introduction

In Chaps. 3 and 7, we have seen several types of models, specified by their quantile functions, that can provide adequate representations of lifetime data. These will be candidate distributions in specific real situations. The selection of one of

them or a new one is dictated by finding out how well it can justify the data generating mechanism and satisfy well other criteria like goodness of fit. Perhaps, the most important requirement in all modelling problems is that the chosen lifetime distribution captures the failure patterns that are inherent in the empirical data. Often, the features of the failure mechanism are assessed from the data with the aid of the ageing concepts discussed earlier in Chap. 5. For instance, it could be the shape of the hazard or mean residual quantile function, assessed from a plot of the observed failure times. Based on this preliminary knowledge, the choice of the distribution can be limited to one from the corresponding ageing class discussed in Chap. 4. Once the question of a suitable model is resolved as an initial choice, the problem then is to test its adequacy against the observed failure times. This is accomplished by first estimating the parameters of the distribution and then carrying out a goodness-of-fit test. Alternatively, nonparametric methods can also be employed to infer various reliability characteristics. In this chapter, we address both general parametric methods and nonparametric procedures from a quantile-based perspective.

Our basic objective in estimation is to find estimates that are as close as possible to the true values of the model parameters. There are different criteria which ensure proximity between the estimate and the true parameter value and accordingly different approaches can be prescribed that meet the desired criteria. One method is to seek estimates by matching the basic characteristics of the chosen model with those in the sample. This includes the method of percentiles, method of moments involving conventional moments, L -moments and probability weighted moments, and then identifying basic characteristics such as location, dispersion, skewness, kurtosis and tail behaviour. A second category of estimation procedures are governed by optimality conditions that renders the difference between the fitted model and the observed data as small as possible or that provides estimates which are most probable. In the following sections, we describe various methods of estimation as well as the properties of these estimates.

9.2 Method of Percentiles

Recall from Chap. 1 that the p th percentile in a set of observations is the value that has $100p\%$ of values below it and $100(1-p)\%$ values above it. Let X_1, X_2, \dots, X_n be a random sample from a population with distribution function $F(x; \Theta)$, or equivalently quantile function $Q(u; \Theta)$, where Θ is a vector of parameters consisting of one or more elements. The sample observations are arranged in order of magnitude with $X_{r:n}$ being the r th order statistic. Then, the sample (empirical) distribution function is defined as

$$F_n(x) = \begin{cases} 0, & x \leq X_{1:n} \\ \frac{i}{n}, & X_{i-1:n} < x \leq X_{i:n} \text{ for } i = 1, 2, \dots, n-1 \\ 1, & x \geq X_{n:n}. \end{cases}$$

Obviously, $F_n(x)$ is the fraction of the sample observations that does not exceed x . The empirical (sample) quantile function then becomes

$$Q_n(p) = F_n^{-1}(p) = \inf\{x | F_n(x) \geq p\} \tag{9.1}$$

which is a step function with jump $\frac{1}{n}$. Notice that (9.1) can be interpreted as a function ξ_p such that the number of observations $\leq \xi_p$ is $\geq [np]$ and the number of observations $\geq \xi_p$ is $\geq [n(1 - p)]$. Thus, e.g.,

$$\begin{aligned} \xi_p &= X_{[np]+1:n} \text{ if } np \text{ is not an integer} \\ &= X_{[np]:n} \text{ if } np \text{ is an integer.} \end{aligned}$$

In practice, some of the other methods of calculating ξ_p are as follows:

- (i) Set $p(n + 1) = k + a$, n being the sample size, k an integer, and $0 \leq a < 1$. Then,

$$\xi_p = \begin{cases} X_{k:n} + a(X_{k+1:n} - X_{k:n}), & 0 < k < n \\ X_{1:n}, & k = 0 \\ X_{n:n}, & k = n \end{cases}$$

(see Sect. 3.2.1);

- (ii) In some software packages, the setting is $1 + p(n - 1) = k + a$;
- (iii) Calculate np . If it is not an integer, round it up to the next higher integer k and $X_{k:n}$ is the value. If np is an integer k , take

$$\xi_p = \frac{1}{2}[X_{k:n} + X_{k+1:n}].$$

The value $X_{[np]+1:n}$ is popular as it assures the monotonic nature of ξ_p in the sense that if x is the p -quantile and y is the q -quantile with $p < q$, then $y < x$.

Some properties of ξ_p as an estimate of $Q(p)$ are described in the following theorems.

Theorem 9.1. *If there is a unique value of $Q(p)$ satisfying*

$$P(X \leq Q(p)) \geq p \text{ and } P(X \geq Q(p)) \geq 1 - p,$$

then $\xi_p \rightarrow Q(p)$ as $n \rightarrow \infty$ with probability 1.

Theorem 9.2. *Let $F(x)$ have a density $f(x)$ which is continuous. If $Q(p)$ is unique and $f(Q(p)) > 0$, then*

- (i) $\sqrt{n}(\xi_p - Q(p)) = n^{\frac{1}{2}}\{p - F_n(Q(p))\}[f(Q(p))]^{-1} + O(n^{-\frac{1}{4}}(\log n)^{\frac{3}{4}})$;
- (ii) $\sqrt{n}(\xi_p - Q(p))$ is asymptotically distributed as

$$N\left(0, \frac{p(1-p)}{n[f(Q(p))]^2}\right).$$

In particular, the asymptotic distribution of the sample median is normal with mean as the population median and variance $\frac{[f(M)]^{-2}}{4n}$, where $M = Q(\frac{1}{2})$ is the population median. For proofs of the above theorems and further results on the asymptotic behaviour of ξ_p , one may refer to Bahadur [44], Kiefer [324], Serfling [527] and Csorgo and Csorgo [160]. Sometimes, the following bound may be useful in evaluating the bias involved in estimating $Q(p)$ by ξ_p . For $0 < p < 1$ and unique $Q(p)$, for all n and every $\varepsilon > 0$,

$$P[|\xi_p - Q(p)| > \varepsilon] \leq 2 \exp[-2n\delta^2],$$

where $\delta = \min(F(Q(p) + \varepsilon) - p, p - F(Q(p) - \varepsilon))$.

A multivariate generalization of Theorem 9.2 is as follows; see Serfling [527].

Theorem 9.3. *Let $0 < p_1 < \dots < p_k < 1$. Suppose F has a density f in the neighbourhood of $Q(p_1), \dots, Q(p_k)$ and f is positive and continuous at $Q(p_1) \dots Q(p_k)$. Then, $(\xi_{p_1}, \xi_{p_2}, \dots, \xi_{p_k})$ is asymptotically normal with mean vector $(Q(p_1), \dots, Q(p_k))$ and covariance $\frac{1}{n}\sigma_{ij}$, where*

$$\sigma_{ij} = \frac{p_i(1-p_j)}{f(Q(p_i))f(Q(p_j))}, \quad i \leq j,$$

and $\sigma_{ij} = \sigma_{ji}$ for $i > j$.

Since the order statistic $X_{k:n}$ is equivalent to the sample distribution function $F_n(x)$, the sample quantile may be expressed as

$$\xi_p = \begin{cases} X_{[np]:n}, & np \text{ is an integer} \\ X_{[np]+1:n}, & np \text{ is not an integer.} \end{cases}$$

By inverting this relation, we get

$$X_{k:n} = \xi_{\frac{k}{n}}, \quad 1 \leq k \leq n,$$

and so any discussion of order statistics could be carried out in terms of sample quantiles and vice versa.

Bahadur [44] has given representations for sample quantiles and order statistics. His results with subsequent modifications are of the following form:

1. If F is twice differentiable at $Q(p)$, $0 < p < 1$, with $q(p) > 0$, then

$$\xi_p = Q(p) + [p - F_n(Q(p))]q(p) + R_n,$$

where, with probability 1,

$$R_n = O(n^{-\frac{3}{4}}(\log n)^{\frac{3}{4}}), \quad n \rightarrow \infty.$$

Alternatively,

$$n^{\frac{1}{2}}(\xi_p - Q(p)) = n^{\frac{1}{2}}\{p - F_n(Q(p))\}q(p) + O(n^{-\frac{1}{4}}(\log n)^{\frac{3}{4}}),$$

$n \rightarrow \infty$;

2. $n^{\frac{1}{2}}(\xi_p - Q(p))$ and $n^{\frac{1}{2}}[p - F_n(Q(p))]q(p)$ each converge in distribution to $N(0, p(1-p)q^2(p))$;
3. Writing $Y_i = Q(p) + [p - I(X_i \leq Q(p))]q(p)$, $i = 1, 2, \dots$,

$$\xi_p = \frac{1}{n} \sum_{i=1}^n Y_i + O(n^{-\frac{3}{4}}(\log n)^{\frac{3}{4}}), \quad n \rightarrow \infty,$$

or with probability 1. ξ_p is asymptotically the mean of the first n values of Y_i . Thus, we have a representation of the sample quantile as a sample mean. Consider a sequence of order statistics $X_{k_n:n}$ for which $\frac{k_n}{n}$ has a limit. Provided

$$\frac{k_n}{n} = p + \frac{k}{n^{\frac{1}{2}}} O\left(\frac{1}{n^{\frac{1}{2}}}\right), \quad n \rightarrow \infty,$$

$n^{\frac{1}{2}}(X_{k_n:n} - \xi_p)$ converge to $kq(p)$ with probability 1, and $n^{\frac{1}{2}}(X_{k_n:n} - Q(p))$ converge in distribution to $N(kq(p), p(1-p)q^2(p))$.

In other words, $X_{k_n:n}$ and ξ_p are roughly equivalent as estimates of $Q(p)$. Rojo [510] considered the problem of estimation of a quantile function when it is more dispersed than distribution function, based on complete and censored samples. Rojo [511] subsequently developed an estimator of quantile function under the assumption that the survival function is increasing hazard rate on the average (IHRA). The estimator of the quantile function in the censored sample case is also given. He has shown that estimators of $Q(u)$ are uniformly strongly consistent.

The percentiles of the population are, in general, functions of the parameter Θ in the model. In the percentile method of estimation, we choose as many percentile points as there are model parameters. Equating these percentile points of the population with the corresponding sample percentiles and solving the resulting equations, we obtain the estimate of Θ . This method ensures that the model fits exactly at the specific points chosen. Since the method does not specify which percentiles are to be chosen in a practical situation, some judgement is necessary in the choice of the percentile points. Issues such as the interpretation of the model parameters and the purpose for which the model is constructed could be some of the guidelines. Shapiro and Gross [535] pointed out that it is a good practice to choose percentiles where inferences are drawn and not to estimate a percentile where interpolation is required between two highest or two lowest values. As a general

practice, they recommended using $p = 0.05$ and $p = 0.95$ for moderate samples and $p = 0.01$ and $p = 0.99$ for somewhat larger samples. In the case of two parameters, one of the above two sets, when there are three parameters augment these by the median $p = 0.50$, and when there are four parameters, the two extreme points along with the quartiles $p = 0.25$ and $p = 0.75$ are their recommendations. The percentile estimators are generally biased, less sensitive to the outliers, and may not guarantee that the mean and variance of the approximating distribution correspond to the sample values.

Example 9.1. Suppose we have a sample of 100 observations from the loglogistic distribution with

$$Q(u) = \frac{1}{\alpha} \left(\frac{u}{1-u} \right)^{\frac{1}{\beta}}, \quad \alpha, \beta > 0.$$

Choosing the values $p = 0.05$ and $p = 0.95$ for matching the population and sample quantiles, we look at the order statistics $X_{5:100}$ and $X_{95:100}$. Then, we have the equations

$$\frac{1}{\alpha} \left(\frac{0.05}{0.95} \right)^{\frac{1}{\beta}} = X_{5:100}, \quad (9.2)$$

$$\frac{1}{\alpha} \left(\frac{0.95}{0.05} \right)^{\frac{1}{\beta}} = X_{95:100}, \quad (9.3)$$

yielding

$$\frac{1}{\alpha^2} = X_{5:100} \times X_{95:100}$$

or $\hat{\alpha} = [X_{5:100} \times X_{95:100}]^{-\frac{1}{2}}.$

Upon substituting $\hat{\alpha}$ in either (9.2) or (9.3) and solving for β , we obtain the estimate of β as

$$\hat{\beta} = \left(\frac{\log 19}{\log \hat{\alpha} X_{5:100}} \right).$$

Instead of using percentiles as such, various quantile-based descriptors of the distribution such as median (M), interquantile range (IQR), measures of skewness (S) and Kurtosis (T), mentioned earlier in Sect. 1.4, may also be matched with the corresponding measures in the sample. The idea is that the fitted distribution has approximately the same distributional characteristics as the observed one. The number of equations should be the same as the number of parameters and the characteristics are so chosen that all the parameters are represented. If there is a parameter representing location (scale), it is a good idea to equate it to the median M (IQR). Some results concerning the asymptotic distributions involving the statistics

are relevant in this context. See Sects. 1.4 and 1.5 for the definitions of various measures and Chap. 3 in which percentile method is applied for various quantile functions.

Theorem 9.4. *The sample interquartile range*

$$iqr = \frac{1}{2}(\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}})$$

is asymptotically normal

$$N\left(\frac{1}{2}\left(Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)\right), \frac{1}{64n}\left(\frac{3}{f(Q(\frac{3}{4}))} - \frac{2}{f(Q(\frac{1}{4}))f(Q(\frac{3}{4}))} + \frac{3}{f^2(Q(\frac{1}{4}))}\right)\right).$$

Note that $IQR = \frac{1}{2}\left(Q(\frac{3}{4}) - Q(\frac{1}{4})\right)$, and that iqr is strongly consistent for IQR.

Theorem 9.5. *The sample skewness s and the sample Moor's kurtosis t possess the following properties:*

$$(s, t) \text{ is consistent for } (S, T)$$

and

$$\sqrt{n}(s - S, t - T)^* \text{ has asymptotic bivariate normal distribution}$$

with mean $(0, 0)^*$ and dispersion matrix $\phi'(c)A(\phi'(c))^*$, where

$$A = (\sigma_{ij}), \quad \sigma_{ij} = \frac{i(8-j)}{64f(E_i)f(E_j)}, \quad i \leq j,$$

$$\phi(c) = (S, T)^*, \quad E_i = \frac{i}{8}, \quad i, j = 1, 3, 5, 7.$$

and $*$ denotes the transpose.

Some other nonparametric estimators of $Q(u)$ have been suggested in literature. Kaigh and Lachenbruch [308] suggested consideration of a subsample of size k without replacement from a complete sample of size n . Then, by defining the total sample estimator of $Q(u)$ as the average of all possible subsamples of size k , they arrived at

$$\hat{Q}_1(p) = \sum_{r=1}^k w_r X_{r:n},$$

where

$$w_r = \frac{\binom{r-1}{j-1} \binom{n-r}{k-j}}{\binom{n}{k}} \text{ and } j = (k+1)p.$$

The choice of k is such that the extreme order statistics should have negligible weight. A somewhat different estimator is proposed in Harrell and Davis [262] of the form

$$\hat{Q}_2(p) = \sum_{r=1}^n w_r X_{r:n},$$

where

$$w_r = I_{\frac{r}{n}}(p(n+1), (1-p)(n+1)) - I_{\frac{r-1}{n}}(p(n+1), (1-p)(n+1))$$

with $I_x(a, b)$ being the incomplete beta function. Both $\hat{Q}_1(p)$ and $\hat{Q}_2(p)$ have asymptotic normal distribution. Specific cases of estimation of the exponential and Weibull quantile functions have been discussed by Lawless [378] and Mann and Fertig [410].

9.3 Method of Moments

The method of moments is also a procedure that matches the sample and population characteristics. We consider three such characteristics here, viz., the conventional moments, L moments and probability weighted moments.

9.3.1 Conventional Moments

In this case, either the raw moments $\mu'_r = E(X^r)$ or the central moments $\mu_r = E(X - \mu)^r$ are equated to the same type of sample moments. When μ'_r is used, we construct the equations

$$\mu'_r = \frac{1}{n} \sum_{i=1}^n X_i^r = m'_r, \quad r = 1, 2, 3, \dots$$

where X_1, X_2, \dots, X_n are independent and identically distributed and the number of such equations is the same as the number of parameters in the distribution. The sample moments

$$m'_r = \int_0^{\infty} x^r dF_n(x)$$

as the estimate of μ'_r have the following properties:

- (i) m'_r is strongly consistent for μ'_r ;
- (ii) $E(m'_r) = \mu'_r$, i.e., the estimates are unbiased;
- (iii) $V(m'_r) = \frac{\mu'_{2r} - \mu'^2_r}{n}$;
- (iv) If $\mu_{2r}' < \infty$, the random vector $n^{-1}(m'_1 - \mu'_1, m'_2 - \mu'_2, \dots, m'_n - \mu'_n)$ converges in distribution to a n -variate normal distribution with mean vector $(0, 0, \dots, 0)$ and covariance matrix $[\sigma_{ij}]$, $i, j = 1, 2, \dots, n$, where $\sigma_{ij} = \mu'_{i+j} - \mu'_i \mu'_j$.

On the other hand, if we use the central moments, the equations to be considered become

$$\mu_r = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r = m_r, \quad \bar{X} = m'_1.$$

The statistic m_r estimates μ_r with the following properties:

- (a) m_r is strongly consistent for μ_r ;
- (b) $E(m_r) = \mu_r + \frac{\frac{1}{2}r(r-1)\mu_r - 2\mu_2 - r\mu_r}{n} + O(n^{-2})$;
- (c) $V(m_r) = \frac{1}{n}(\mu_{2r} - \mu_r^2 - 2r\mu_{r-1}\mu_{r+1} + r^2\mu_2\mu_{r-1}^2) + O(\frac{1}{n^2})$;
- (d) If $\mu_{2r} < \infty$, the random vector $n^{\frac{1}{2}}(m_2 - \mu_2, \dots, m_r - \mu_r)$ converges in distribution to $(r - 1)$ dimensional normal distribution with mean vector $(0, 0, \dots, 0)$ and covariance matrix $[\sigma_{ij}]$, $i, j = 2, 3, \dots, r$, where

$$\sigma_{ij} = \mu_{i+j+2} - \mu_{i+1}\mu_{j+1} - (i+1)\mu_i\mu_{j+2} - (j+1)\mu_{i+2}\mu_j + (i+1)(j+1)\mu_i\mu_j\mu_2.$$

In general, m_r gives biased estimator of μ_r . The correction factor required to make them unbiased and the corresponding statistics are

$$M_2 = \frac{n}{n-1}m_2,$$

$$M_3 = \frac{n}{(n-1)(n-2)}m_3,$$

$$M_4 = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)}m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)}m_2^2.$$

Occasionally, the parameters are also estimated by matching μ'_1, μ_2, β_1 and β_2 with the corresponding sample values. For example, the estimation of parameters of the lambda distributions is often done in this manner.

Example 9.2. Consider the generalized Pareto model (Table 1.1) with

$$Q(u) = \frac{b}{a} [(1-u)^{-\frac{a}{a+1}} - 1].$$

In this case, the first two moments are

$$\mu'_1 = \int_0^1 Q(p) dp = b$$

and

$$\mu'_2 = \int_0^1 Q^2(p) dp = \frac{2b^2}{1-a}.$$

Hence, we form the equations

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

$$\frac{2b^2}{1-a} = \frac{1}{n} \sum X_i^2,$$

and solve them to obtain the moment estimates of a and b as

$$\hat{a} = 1 - \frac{2n\bar{X}^2}{\sum X_i^2} \quad \text{and} \quad \hat{b} = \bar{X}.$$

9.3.2 *L-Moments*

In the method of L -moments, the logic is the same as in the case of usual moments except that we equate the population L -moments with those of the sample and then solve for the parameters. Here again, the number of equations to be considered is the same as the number of parameters to be estimated. Thus, we consider r equations

$$L_r = l_r, \quad r = 1, 2, \dots,$$

where

$$L_r = \int_0^1 \sum_{k=0}^{r-1} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} u^k Q(u) du$$

and

$$l_r = \frac{1}{n} \sum_{j=0}^{r-1} p_{rj} \left(\sum_{i=1}^n \frac{(r-1)_{(j)}}{(n-1)_{(j)}} \right) \quad (9.4)$$

with

$$p_{ij} = \frac{(-1)^{i-1-j} (i+j-1)!}{(j!)^2 (i-j-1)!},$$

when the model to be fitted contains r parameters. The expressions for the first four L -moments L_1 through L_4 are given in (1.34)–(1.37) (or equivalently (1.38)–(1.41)). Next, we have the sample counterparts as

$$\begin{aligned}
 l_1 &= \frac{1}{n} \sum_{i=1}^n X_{(i)} = \bar{X}, \\
 l_2 &= \frac{1}{n(n-1)} \sum_{i=1}^n (2i-1-n)X_{i:n}, \\
 l_3 &= \frac{1}{n(n-1)(n_2)} \sum_{i=1}^n \{6(i-1)(i-2) - 6(i-1)(n-2) \\
 &\quad + (n-1)(n-2)\} X_{i:n}, \\
 l_4 &= \frac{1}{n(n-1)(n_2)(n-3)} \sum_{i=1}^n \{20(i-1)(i-2)(i-3) - 30(i-1)(i-2)(n-3) \\
 &\quad + 12(i-1)(n-2)(n-3) - (n-1)(n-2)(n-3)\} X_{i:n}.
 \end{aligned}$$

Regarding properties of l_r as estimates of L_r , we note that l_r is unbiased, consistent and asymptotically normal (Hosking [276]). Elamir and Seheult [187] have obtained expressions for the exact variances of the sample L -moments. They have used an equivalent representation of (9.4) in the form

$$l_r = \sum_{k=0}^{r-1} p_{r-1,k}^* b_k, \tag{9.5}$$

where

$$p_{r-1,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}$$

and

$$b_k = \frac{1}{n_{(k+1)}} \sum_{i=1}^n (i-1)_{(k)} X_{i:n}.$$

For a sample of size n , (9.5) is also expressible in the vector form

$$l = bC^T$$

with $l = (l_1, l_2, \dots, l_n)$ and $b = (b_0, b_1, \dots, b_{k-1})$ and C is a triangular matrix with entries $p_{r-1,k}^*$. So,

$$V(l) = C\Theta C^T, \quad \Theta = V(b). \tag{9.6}$$

As special cases, we have

$$V(l_1) = \frac{\sigma^2}{n}, \tag{9.7}$$

where σ^2 is the population variance and

$$\begin{aligned} V(l_2) = \frac{1}{n(n-1)} & \left[\frac{4(n-2)}{3} \{E(Y_{3:3}^2) + E(Y_{1:3}Y_{2:3}) + E(Y_{2:3}Y_{3:3})\} \right. \\ & - 2(n-3)E(Y_{1:2}Y_{2:2}) - 2(n-2)E(Y_{2:2}^2) + (n-1)E(Y_{1:1}^2) \\ & \left. - 2(2n-3)E(Y_{2:2}^2) + E(Y_{1:1})\{4(2n-3)E(Y_{2:2}) - 5(n-1)E(Y_{1:1})\} \right]. \end{aligned} \tag{9.8}$$

In the case of the first four sample moments,

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ -1 & -6 & 6 & 0 \\ -1 & 12 & -30 & 20 \end{pmatrix}$$

along with

$$\begin{aligned} V(l_1) &= \theta_{00}, \\ V(l_2) &= 4\theta_{11} - 4\theta_{01} + \theta_{00}, \end{aligned}$$

were used to derive the expressions in (9.7) and (9.8). Furthermore,

$$\text{Cov}(l_1, l_2) = \frac{1}{3n} [E(Y_{3:3} - Y_{2:3})^2 - E(Y_{2:3} - Y_{1:3})^2]$$

and

$$\begin{aligned} \text{Cov}(l_1, l_r) &= \frac{1}{n} \left[\int_0^1 u^2 P_{r-1}^*(u) du - \int_0^1 \int_0^u [uv F(P_{r-1}^*(u))]’ dudv \right. \\ & \left. - (-1)^r \int_0^1 \int_0^v uv [(1-G)P_{r-1}^*(u)(1-G)]’ dudv, \right. \end{aligned}$$

where $G(y) = v$ and $F(x) = u$. Also, if we define

$$\theta_{kl} = \text{Cov}(b_k, b_l), \tag{9.9}$$

then

$$\hat{\theta}_{kl} = b_k b_l - \frac{1}{n(n+l+2)} \sum_{1 \leq i < j \leq n} [(i-1)_{(k)}(j-k-2)_{(l)} + (i-1)_{(l)}(j-l-2)_{(k)}] X_{i:n} X_{j:n}$$

is a distribution-free unbiased estimator of (9.9). In the above discussion, $Y_{r:n}$ denotes the conceptual order statistics of the population. The expression for $V(l_2)$ is equivalent to the estimate of Nair [455]. In finding the variance of the ratio of two sample L -moments, the approximation

$$V\left(\frac{X}{Y}\right) \doteq \left[\frac{V(X)}{E(X)^2} + \frac{V(Y)}{E(Y)^2} - \frac{2\text{Cov}(X,Y)}{E(X)E(Y)} \right] \left(\frac{E(X)}{E(Y)}\right)^2$$

is useful. Thus, approximate variances of the sample L -skewness and kurtosis can be obtained. The sample L -moment ratios are consistent but not unbiased.

A more general sampling scheme involving censoring of observations has been discussed recently. Let T_1, T_2, \dots, T_n be independent and identically distributed lifetimes following distribution function $F(x)$. Assume that lifetimes are censored on the right by independent and identically distributed random variables Y_1, Y_2, \dots, Y_n having common distribution function $H(x)$. Further, let Y_i 's be independent of the T_i 's. Thus, we observe only the right censored data of the form $X_i = \min(T_i, Y_i)$. Define indicator variables

$$\Delta_i = \begin{cases} 1 & \text{if } T_i \leq Y_i \\ 0 & \text{if } T_i > Y_i \end{cases}$$

so that $\Delta_i = 1$ indicates T_i is uncensored and $\Delta_i = 0$ indicates T_i is censored. The distribution of each X_i is then

$$G(x) = 1 - (1 - F(x))(1 - H(x)).$$

To estimate the distribution function of the censored samples (X_i, Δ_i) , the Kaplan–Meir [311] product limit estimator is popular.

The estimator of the survival function is

$$S_n(t) = \prod_{j: X_{(j)} \leq t} \left(\frac{n-j}{n-j+1} \right)^{\Delta(j)}, \quad t \leq X_{(n)},$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ are ordered X_i 's and $\Delta(j)$ is the censoring status corresponding to $X_{(j)}$.

The estimator of the r th L -moment for right censored data (Wang et al. [576]) is

$$\hat{L}_r = \sum_{j=1}^n X_{(j)} U_{j(r)},$$

where

$$U_{j(r)} = \sum_{k=0}^{j-1} r^{-1} \binom{r-1}{k} [B_{p,q}(1 - S_n(X_{(j)})) - B_{p,q}(1 - S_n(X_{(j-1)}))]$$

with $X_{(0)} = 0$, $p = r - k$ and $q = k + 1$.

Now, let $T = \min(X, C)$, where X denotes the failure time and C denotes the noninformative censoring time. For a constant $0 \leq u_0 < 1$ and $Q(u_0) < T^*$, where T^* is the minimum of the upper most support points of the failure time X and the censoring time C , suppose that $F''(x)$ is bounded on $[0, Q(u_0) + \Delta]$, $\Delta > 0$ and $\inf_{0 \leq u \leq u_0} f(Q(u)) > 0$. Then (Cheng [145]),

(i) with probability one

$$\sup_{0 \leq u \leq u_0} |\hat{Q}(u) - Q(u)| = O(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}}),$$

(ii)

$$\sup_{0 \leq u \leq u_0} |n^{\frac{1}{2}} f(Q(u))(\hat{Q}(u) - Q(u) - G_n(u))| = O(n^{-\frac{1}{3}}(\log n)^{\frac{3}{2}}),$$

where $G_n(u)$ is a sequence of identically distributed Gaussian process with zero mean and covariance function

$$\text{Cov}(G_n(u_1), G_n(u_2)) = (1 - u_1)(1 - u_2) \int_0^n \frac{dt}{(1 - t)^2 [1 - H(F^{-1}(t))]},$$

with $u_1 \leq u_2$, $H(x) = 1 - (1 - F(x)(1 - G(x)))$ and $G(x)$ is the distribution function of the censoring time C .

Under the above regularity conditions, Wang et al. [576] have shown that, as $n \rightarrow \infty$,

- (i) $\hat{L}_r = L_r + o(n^{-\frac{1}{2}}(\log \log n)^{\frac{1}{2}})$;
- (ii) $n^{\frac{1}{2}}(\hat{L}_r - L_r)$, $r = 1, 2, \dots, n$ converges in distribution to multivariate normal $(0, \Sigma)$, where Σ has its elements as

$$\Sigma_{rs} = \iint_{x \leq y} \frac{P_{r-1}^*(x)P_{s-1}^*(y) + P_{s-1}^*(x)P_{r-1}^*(y)}{f(Q(x))f(Q(y))} \text{Cov}(G_n(x), G_n(y)) dx dy,$$

with

$$P_{r-1}^*(u) = \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r+k-1}{k} u^k$$

is the $(r - 1)$ th shifted Legendre polynomial;

(iii) the vector

$$n^{\frac{1}{2}}(\hat{L}_1 - L_1, \hat{L}_2 - L_2, \hat{\tau}_3 - \tau_3, (\hat{\tau}_3 - \tau_3) \dots (\hat{\tau}_m - \tau_m))$$

converges in distribution to multivariate normal $(0, \Lambda)$, where Λ has its elements as

$$\Lambda_{rs} = \begin{cases} \Sigma_{rs}, & r \leq 2, s \leq 2 \\ \frac{(\Sigma_{rs} - \tau_r \Sigma_{2s})}{L_2}, & r \geq 2, s \leq 2 \\ \frac{\Sigma_{rs} - \tau_r \Sigma_{2s} - \tau_s \Sigma_{2r} + \tau_r \tau_s \Sigma_{22}}{L_2^2}, & r \geq 3, s \geq 3. \end{cases}$$

There are several papers that deal with L -moments of specific distributions and comparison of the method of L -moments with other methods of estimation. Reference may be made, e.g., to Hosking [277], Pearson [488], Guttman [256], Gingras and Adamowski [217], Hosking [278], Sankarasubramonian and Sreenivasan [517], Chadjiconstantinidis and Antzoulakos [131], Hosking [280], Karvanen [312], Ciumara [150], Abdul-Moniem [3], Asquith [40] and Delicade and Gorja [169]. Illustration of the method of L -moments with some real data can be seen in Sect. 3.6 for different models.

9.3.3 Probability Weighted Moments

Earlier in Sect. 1.4, we defined the probability weighted moment (PVM) of order (p, r, s) as

$$M_{p,r,s} = E(X^p F^r(X) \bar{F}^s(X))$$

which is the same as

$$\begin{aligned} M_{p,r,s} &= \int_0^\infty x^p F^r(x) \bar{F}^s(x) f(x) dx \\ &= \int_0^1 [Q(u)]^p u^r (1-u)^s du, \end{aligned}$$

provided that $E(|X|^p) < \infty$. Commonly used quantities are

$$\beta_r = \int_0^1 [Q(u)] u^r du$$

and

$$\alpha_r = \int_0^1 [Q(u)] (1-u)^r du,$$

which are the special cases $M(1, r, 0)$ and $M(1, 0, r)$. Since

$$\alpha_r = \sum_{s=0}^r \binom{r}{s} (-1)^s \beta_s \text{ and } \beta_r = \sum_{s=0}^r \binom{r}{s} (-1)^s \alpha_s,$$

characterization of a distribution with finite mean by α_r or β_r is interchangeable. A natural estimate of α_r (also called the nonparametric maximum likelihood estimate) based on ordered observations $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ is

$$\hat{\alpha}_r = \int_0^\alpha x(1 - F_n(x))^r dF_n(x) = \frac{1}{n} \sum_{i=1}^n X_{i:n} \left(1 - \frac{i}{n}\right)^r.$$

The asymptotic covariance of the estimator is

$$\begin{aligned} \sigma_{rs} &= \text{Cov}(\hat{\alpha}_r, \hat{\alpha}_s) \\ &= \frac{1}{n} \iint_{x < y} [1 - F(x)]^r [1 - F(y)]^s F(x) (1 - F(x)) dx dy. \end{aligned}$$

Similarly, the estimate of β_r is

$$\hat{\beta}_r = \int_0^\infty x [F_n(x)]^r dF_n(x) = \frac{1}{n} \sum_{i=1}^n X_{i:n} \left(\frac{i}{n}\right)^r,$$

with asymptotic covariance

$$\text{Cov}(\hat{\beta}_r, \hat{\beta}_s) = \frac{1}{n} \iint_{x < y} [F(x)]^r [F(y)]^s F(x) [1 - F(y)] dy.$$

Landwehr and Matalas [373] have shown that

$$b_r = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} X_{i:n} \tag{9.10}$$

is an unbiased estimator of β_r . Similarly, for α_r , we have the estimator

$$a_r = \frac{1}{n} \sum_{i=1}^n \frac{(n-i)(n-i-1)\dots(n-i+r-1)}{(n-1)(n-2)\dots(n-r)}.$$

Hosking [278] has developed estimates based on censored samples. In Type I censoring from a sample of size n , m are observed and $(n - m)$ are censored above a known threshold T so that m is a random variable with binomial distribution. The estimate based on the uncensored sample of m values is (9.10) with m replacing n and when the $n - m$ censored values are replaced by T , the estimate is given by

$$\bar{b}_r = \frac{1}{n} \left[\sum_{j=1}^m \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} X_{i:n} + \sum_{i=m+1}^m \frac{(i-1)\dots(i-r)}{(n-1)\dots(n-r)} T \right].$$

Assume that $F(t) = v$. Conditioned on the achieved value of m , the uncensored values are a random sample of size m from the distribution with quantile function $Q(uv)$, $0 < u < 1$. The population PWM's for this distribution are

$$\beta_{r,m} = \frac{1}{v^{r+1}} \int_0^v u^r Q(u) du.$$

On the other hand, the completed sample is of size n from the distribution with quantile function

$$Q_1(u) = \begin{cases} Q(u), & 0 < u < v \\ Q(v), & v \leq u < 1 \end{cases}$$

and hence its PWM's are

$$\begin{aligned} \beta_{r,n} &= \int_0^1 u^r Q_1(u) du \\ &= \int_0^v u^r Q(u) du + \frac{1 - v^{r+1}}{r+1} Q(v). \end{aligned}$$

The asymptotic distributions in this case are derived as in the case of the usual PWM's. Furrer and Naveau [205] have examined the small-sample properties of probability weighted moments.

As in the case of the other two moments, we equate the sample and population probability weighted moments for the estimation of the parameters. In such cases, it is useful to adopt the formulas

$$\bar{b}_r = \frac{1}{n} \sum_{i=1}^n x_{(i)} p_{(i)}^r$$

and

$$\bar{a}_r = \frac{1}{n} \sum_{i=1}^n x_{(i)} (1 - p_{(i)})$$

with $p_{(i)}$ as some ordered suitably chosen probabilities. Gilchrist [215] has prescribed the choice for $p_{(i)}$ as $\frac{i}{n+1}$, or inverse of the beta function $(0.5, i, n - i + 1)$, or $p_i = \frac{i-0.5}{n}$.

Example 9.3. Consider the Govindarajulu distribution with

$$\begin{aligned} Q(u) &= \sigma((\beta + 1)u^\beta - \beta u^{\beta+1}) \\ \beta_r &= \sigma \int_0^1 [(\beta + 1)u^\beta - \beta u^{\beta+1}] u^r du, \\ &= \frac{\sigma(2\beta + r + 2)}{(\beta + r + 1)(\beta + r + 2)}. \end{aligned}$$

Since we have only two parameters in this case to estimate, the equations to be considered are

$$\beta_r = \bar{b}_r, \quad r = 0, 1,$$

or

$$\frac{2\sigma}{\beta + 2} = \bar{X} = \bar{b}_0$$

and

$$\frac{\sigma(2\beta + 3)}{(\beta + 2)(\beta + 3)} = \frac{1}{n(n+1)} \sum iX_{(i)} = \bar{b}_1.$$

Upon solving these equations, we obtain the estimates

$$\hat{\beta} = \frac{3\bar{b}_0 - 6\bar{b}_1}{2(\bar{b}_1 - \bar{b}_0)}, \quad \hat{\sigma} = \frac{\bar{b}_0(\bar{b}_0 + 2\bar{b}_1)}{4(\bar{b}_0 - \bar{b}_1)}.$$

9.4 Method of Maximum Likelihood

We proceed by writing the likelihood function based on a random sample x_1, \dots, x_n as

$$L(\theta) = f(x_1; \theta)f(x_2; \theta) \dots f(x_n; \theta).$$

Taking $x_i = Q(u_i; \theta)$, we then have

$$\begin{aligned} L(\theta) &= f(Q(u_1, \theta))f(Q(u_2, \theta)) \dots f(Q(u_n, \theta)) \\ &= [q(u_1; \theta)q(u_2; \theta) \dots q(u_n; \theta)]^{-1}. \end{aligned}$$

The estimate of θ is the solution that maximizes $L(\theta)$, or equivalently, we have to solve the equation

$$\frac{d \log L}{d \theta} = - \frac{\sum_i d \log q(u_i; \theta)}{d \theta} = 0$$

for θ . Notice that, in practice, the calculation of $L(\theta)$ requires the derivation of u_i from the equation $x_i = Q(u_i; \theta)$. If $u_i = F(x_i; \theta)$ is explicitly available, a direct solution of the u_i 's are possible from the observed x_i in the fit of $F(x)$ after substituting the estimated values for the parameters. Otherwise, one has to use some numerical method to extract u_i . The observations are ordered. When U is a uniform random variable, X and $Q(U)$ have identical distributions. Let $\hat{Q}(u)$ be a fitted quantile function and $x_{(r)} = Q(u_{(r)})$. If ψ_0 is an initial estimate of u for a given x value, using the first two terms of the Taylor expansion, we have

$$Q(\psi) = Q(u_0) + (u - u_0)Q'(u_0)$$

as an approximation. Solving for u , we get

$$u \doteq u_0 + \frac{x - Q(u_0)}{q(u_0)}. \tag{9.11}$$

In practical problems, the initial value could be $u_{(r)} = \frac{r}{n+1}$. With this value, (9.11) is used iteratively until x differs from $Q(u)$ by ϵ , a small pre-set tolerance value, in a trial. Gilchrist [215] has provided a detailed discussion on the subject and an example of the estimation of the parameters of the generalized lambda distribution and layout for the calculations. The properties of the maximum likelihood estimates, though widely known, is included here for the sake of completeness.

Theorem 9.6. *Let X_1, X_2, \dots, X_n be independent and identically distributed with distribution function $F(x; \theta)$ where θ belongs to an open interval Θ in R , satisfying the following conditions:*

(a)

$$\frac{\partial \log f(x; \theta)}{\partial \theta}, \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2}, \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3}$$

exist for all x ;

(b) for each $\theta_0 \in \Theta$, there exist functions $g_i(x)$ in the neighbourhood of θ_0 such that

$$\left| \frac{\partial f}{\partial \theta} \right| \leq g_1(x), \quad \left| \frac{\partial^2 f}{\partial \theta^2} \right| \leq g_2(x), \quad \left| \frac{\partial^3 \log f}{\partial \theta^3} \right| \leq g_3(x)$$

for all x , and

$$\int g_1(x) dx < \infty, \quad \int g_2(x) dx < \infty, \quad E g_3(X) dx < \infty$$

in the neighbourhood of θ_0 ;

(c) $0 < E\left(\frac{\partial \log f(X; \theta)}{\partial \theta}\right)^2 < M < \infty$ for each θ , then with probability 1, the likelihood equations

$$\frac{\partial L}{\partial \theta} = 0$$

admit a sequence of solutions $\{\hat{\theta}_n\}$ with the following properties:

- (i) $\hat{\theta}_n$ is strongly consistent for θ ;
- (ii) $\hat{\theta}_n$ is asymptotically distributed as

$$N\left(\theta, \frac{1}{nE\left(\frac{\partial \log f(X; \theta)}{\partial \theta}\right)^2}\right).$$

When θ contains more than one element, then also the sequence of vector values $(\hat{\theta}_n)$ satisfies consistency and asymptotic normality $(\theta, \frac{1}{nI(\theta)})$ where $I(\theta)$, called the information matrix, has its elements as

$$E\left(\frac{\partial \log f(X; \theta)}{\partial \theta_i}, \frac{\partial \log f(X; \theta)}{\partial \theta_i}\right)$$

where $\theta = (\theta_1, \dots, \theta_K)$ and $I(\theta)$ has order $K \times K$.

9.5 Estimation of the Quantile Density Function

The quantile density function $q(u)$ is a vital component in the definitions of reliability concepts like hazard quantile function, mean residual quantile function and total time on test transforms. Moreover, it appears in the asymptotic variances of different quantile-based statistics. Babu [43] has pointed out that the estimate of the bootstrap variance of the sample median needs consistent estimates of $q(u)$.

Assume that $q(p) \geq 0$. Then,

$$\frac{Q(v) - Q(p)}{v - p} = q(p) + o(1)$$

as $v \rightarrow p$. Thus, to get an approximation for $q(p)$, it is enough to consider $Q(v) - Q(p)$ for $v \geq u$ near u . As $Q(p)$ is not known, it is replaced by $Q_n(p)$. Since all quantiles $\frac{Q_n(v) - Q(p)}{v - p}$ are close to $q(p)$, a linear combination of these two also will be near $q(p)$. With this as the motivating point, Babu [43] has provided the following results.

Let h be a function on the positive real line such that $h(y)e^y$ is a polynomial of degree not exceeding k ($k \geq 2$ is an integer) and

$$\int_0^\infty h(y)y^j dy = \begin{cases} 1, & j = 0, 1 \\ 0, & j = 2, 3, \dots, k. \end{cases}$$

Then,

$$\sigma^2(x) = \int_0^x \int_0^x h(p)h(v) \min(p, v) dp dv$$

and

$$\sigma^2 = \int_0^\infty \int_0^\infty h(p)h(v) \min(p, v) dudv.$$

Defining $J = (p - \varepsilon, p + \varepsilon) \subset (0, 1)$, for independent variables X_1, \dots, X_n with common distribution function $F(x)$, we have the following two results.

Theorem 9.7. *If $F(x)$ is k times continuously differentiable at $Q(p)$ for $p \in J$ such that $f(x)$ at $Q(p)$ is positive and $E(X^2) < \infty$, then uniformly in $x > 0$,*

$$n^{2\beta} E \left(\frac{D(x, n)}{q(p)} - 1 \right)^2 = \sigma^2(x) + o(1) + n^{2\beta} (1 - L(x, n))^2,$$

where

$$D(x, n) = n^\delta \int_0^x [Q_n(p + vn^{-\delta}) - Q_n(p)] h(v) dv,$$

$$L(x, n) = n^\delta b_1^{-1} \int_0^x b(vn^{-\delta}) h(v) dv,$$

$$\delta = (2k - 1)^{-1}, \beta = \frac{1 - \delta}{2}, b_j = \frac{1}{j!} \frac{d^j Q(p)}{dp^j} \text{ at } p \in J, \text{ and } b(p) = \sum_{j=1}^k b_j p^j.$$

Theorem 9.8. *Let $f_i(x) = \int_x^\infty y^i h(y) dy$ and f_j and f_1 do not have common positive roots for any $2 \leq j \leq k - 1$. If the j th derivative of Q at p is non-zero, for $2 \leq j \leq k - 1$, then*

$$n^{2\beta} E \left(\frac{D(x, n)}{q(p)} - 1 \right)^2 \geq \sigma^2 + o(1)$$

and that the equality occurs at $x = \log n$. As a result, $D(\log n, n)$ is an efficient estimator of $q(u)$ in the mean square sense among the class of estimators $\{D(x, n) | x > 0\}$. Also, $n^\beta (D(\log n, n) (\frac{1}{q(p)} - 1))$ is asymptotically distributed as $N(0, \sigma^2)$.

A histogram type estimator of the form

$$A_n(p) = \frac{Q_n(p + \alpha_n) - Q_n(p - \alpha_n)}{2\alpha_n}, \quad \alpha_n > 0,$$

has been discussed by Bloch and Gastwirth [108] and Bofinger [113]. Its asymptotic distribution is presented in the following theorem.

Theorem 9.9 (Falk [193]). *Let $0 < p_1 < \dots < p_r < 1$ and $Q(p)$ be twice differentiable near p_j with bounded second derivative, $j = 1, 2, \dots, r$. Then, if $\alpha_{nj} \rightarrow 0$ and $n\alpha_{jn} \rightarrow \infty$, $j = 1, 2, \dots, r$,*

$$(2n\alpha_{jn})^{\frac{1}{2}} \left\{ A_{jn}(p_j) - \frac{Q(p_j + \alpha_{jn}) - Q(p_j - \alpha_{jn})}{2\alpha_{jn}} \right\}_{j=1}^r,$$

where

$$A_{jn}(p_j) = \frac{Q_n(p_j + \alpha_{jn}) - Q_n(p_j - \alpha_{jn})}{2\alpha_{jn}},$$

converges in distribution to $\prod_{j=1}^r N(0, q^2(p_j))$, with Π denoting the product measure.

In the above result, if we further assume that $n\alpha_{jn}^3 \rightarrow 0$, then $\frac{Q(p_j + \alpha_{jn}) - Q(p_j - \alpha_{jn})}{2\alpha_{jn}}$ can be replaced $q(p_j)$. Moreover, if Q is three times differentiable near p with bounded third derivative which is continuous at p , an optimal bandwidth α_n^* in the sense of mean square is

$$\alpha_n^* = \left\{ \frac{\frac{6}{1} q(p)}{\frac{2^{\frac{1}{2}}}{q^3(p)}} \right\}^{\frac{2}{5}} n^{-\frac{1}{5}}.$$

Falk [193] considered a kernel estimator of $q(p)$ defined by

$$\hat{k}_n(F_n) = \int_0^1 Q_n(x) \alpha_n^{-2} h\left(\frac{p-x}{\alpha_n}\right) dx,$$

where h is a real valued kernel function with bounded support and $\int h(x) dx = 0$. Notice that $k_n(p)$ is a linear combination of order statistics of the form $\sum_{i=1}^n C_{in} X_{i:n}$, where

$$C_{in} = \frac{1}{\alpha_n^2} \int_{\frac{i-1}{n}}^{\frac{i}{n}} h\left(\frac{p-x}{\alpha_n}\right) dx, \quad i = 1, 2, \dots, n.$$

Some key properties of the above kernel estimator are presented in the following theorem.

Theorem 9.10. *Let $0 < p_1 < \dots < p_r < 1$ and Q be twice differentiable near p_j with bounded second derivative, $j = 1, 2, \dots, r$. Then, if h_j has the properties of h given above and $\alpha_{jn} \rightarrow 0$, $n\alpha_{jn}^2 \rightarrow \infty$,*

$$\left[(n\alpha_{jn})^{\frac{1}{2}} (\hat{k}_{jn}(p_j) - k_{jn}(p_j)) \right]_{j=1}^r$$

converges to $\prod_{j=1}^r N(0, q^2(p_j)) \int H_j^2(y) dy$, where \prod is the product measure,

$$H_j(y) = \int_{-\infty}^y h_j(x) dx,$$

$$\hat{k}_{nj}(p_j) = \int_0^1 Q_n(x) \alpha_{jn}^{-2} h_j\left(\frac{p_j - x}{\alpha_{jn}}\right) dx,$$

and

$$\hat{k}_{nj}(p_j) = \int_0^1 Q(x) \alpha_{jn}^{-2} h_j\left(\frac{p - x}{\alpha_{jn}}\right) dx.$$

Using additional conditions $\int xh(x)dx = -1$ and $n\alpha_n^3 \rightarrow 0$, $k_n(p)$ in Theorem 9.4 can be replaced by $q(p)$. Further, if Q is differentiable $(m + 1)$ times with bounded derivatives which are continuous at p , with $n\alpha_n^{3m+1} \rightarrow 0$, the approximate bias of $\hat{k}_n(p)$ becomes

$$k_n(p) - q(p) = o(\alpha_n^m)$$

and the optimal bandwidth that minimizes the mean squared error $E(\hat{k}_n(p) - q(p))^2$ is

$$\alpha_n^* = \frac{(m + 1)! (\int H^2(y) dy)^{\frac{1}{2}} q(p)}{Q^{m+1}(p) \int x^{m+1} h(x) dx}.$$

Mean squared error of the kernel quantile density estimator is compared with that of the estimate $\tilde{g}(u)$ in Jones [305], where $\tilde{g}(u)$ is the reciprocal of the kernel density estimator given by

$$\tilde{g}(u) = \frac{1}{\hat{f}(Q_n(u))}.$$

It is proved that the former estimator is better than the latter one in terms of the mean squared error.

Estimation of $q(p)$ in a more general framework and for different sampling strategies has been discussed by Xiang [590], Zhou and Yip [603], Cheng [143] and Buhamra et al. [123].

9.6 Estimation of the Hazard Quantile Function

The hazard quantile function in reliability analysis, as described in the preceding chapters, plays a key role in describing the patterns of failure and also in the selection of the model. Sankaran and Nair [515] have provided the methodology for the nonparametric estimation of the hazard function, by suggesting two estimators, with one based on the empirical quantile density function and the other based on a kernel density approach. The properties of the kernel-based estimator and comparative study of the two estimators have also provided by them. Recall that the hazard quantile function is defined as

$$H(p) = [(1-p)q(p)]^{-1}, \quad 0 < p < 1.$$

Suppose that the lifetime X is censored by a non-negative random variable Z . We observe (T, Δ) , where $T = \min(X, Z)$ and $\Delta = I(X \leq Z)$, with

$$T(X \leq Z) = \begin{cases} 1, & X \leq Z \\ 0, & X > Z. \end{cases}$$

If $G(x)$ and $L(x)$ are the distribution functions of Z and T , respectively, under the assumption that Z and X are independent, we have

$$1 - L(x) = (1 - F(x))(1 - G(x)).$$

Let (T_i, Δ_i) , $i = 1, 2, \dots, n$, be independent and identically distributed and each (T_i, Δ_i) has the same distribution as (T, Δ) . This framework includes time censored observations if all the Z_i 's are fixed constants, a Type I censoring when all Z_i 's are the same constant, and Type II censoring if $Z_i = X_{r:n}$ for all i . The first estimator proposed by Sankaran and Nair [515] is

$$\hat{H}(p) = \frac{1}{[1 - F_n(Q_n(p))]q_n(p)},$$

where

$$q_n(p) = n(T_{j:n} - T_{j-1:n}), \quad \frac{j-1}{n} \leq p \leq \frac{j}{n},$$

and $T_{0:n} \equiv 0$. From Parzen [486], it follows that $q_n(p)$ is asymptotically exponential with mean $q(p)$. Thus, $q_n(p)$ is not a consistent estimator of $q(p)$ nor $\hat{H}(p)$ is for $H(p)$.

A second estimator has been proposed by considering a real valued function $K(\cdot)$ such that

- (i) $K(x) \geq 0$ for all x and $\int K(x)dx = 1$,
- (ii) $K(x)$ has finite support, i.e., $K(x) = 0$ for $|x| > c$ for some constant $c > 0$,
- (iii) $K(x)$ is symmetric about zero,
- (iv) $K(x)$ satisfies the Lipschitz condition

$$|K(x) - H(y)| \leq M|x - y|$$

for some constant M . Further, let $\{h_n\}$ be a sequence of positive numbers such that $h_n \rightarrow 0$ as $n \rightarrow \infty$. Define a new estimator as

$$H_n(p) = \frac{1}{h_n} \int_0^1 \frac{1}{[1 - F_n(Q_n(t))]q_n(t)} K\left(\frac{t - p}{h_n}\right) dt \tag{9.12}$$

$$= \frac{1}{h_n} \sum_{i=1}^u \frac{1}{[1 - F_n(T_{i:n})]^n [T_{i:n} - T_{i-1:n}]} \int_{S_{i-1:n}}^{S_{i:n}} K\left(\frac{t - p}{h_n}\right) dt, \tag{9.13}$$

where

$$S_{i:n} = \begin{cases} 0, & i = 0 \\ F_n(T_{i:n}), & i = 1, 2, \dots, n - 1 \\ 1, & i = n. \end{cases}$$

When $S_{i:n} - S_{i-1:n}$ is small, by the first mean value theorem, (9.13) is approximately equal to

$$H_n^*(p) = \frac{1}{h_n} \sum_{i=1}^n \frac{S_{i:n} - S_{i-1:n}}{[1 - F_n(T_{i:n})]^n (T_{i:n} - T_{i-1:n})} K\left(\frac{S_{i:n} - p}{h_n}\right).$$

When no censoring is present, $S_{i:n} - S_{i-1:n} = \frac{i}{n}$ for all i . When heavy censoring is present, $S_{i:n} - S_{i-1:n}$ is large for $i = n$ so that $H_i^*(p)$ need not be a good approximation for $H_n(p)$.

When F is continuous and $K(\cdot)$ satisfies Conditions (i)–(iv) given above, the estimator $H_n(p)$ is uniformly strongly consistent and for $0 < p < 1$, $(\sqrt{n}H_n(p) - H(p))$ is asymptotically normal with mean zero and variance

$$\sigma^2(p) = \frac{n}{(h(n))^2} E \left[\int_0^1 Q_n(t) dM'(t, p) + \int_0^1 F_n(Q_n(t)) \frac{M(t, p)}{(1 - t)} q(t) dt \right]^2.$$

A simulation was carried out in order to make a small sample comparison of $H_n(p)$ and $\hat{H}(p)$ in terms of mean squared error. The random censorship model with $F(t) = 1 - e^{-\lambda t}$ was used by varying λ . Observations were censored with the uniform distribution $U(0, 1)$ with probability 0.3, so that 30% of the observations were censored. As a choice of the kernel function, the triangular density

$$K(x) = (1 - |x|)I(|x| \leq 1)$$

was used. The ratios of the mean squared error of $\hat{H}(p)$ to that of $H_n(p)$ were compared in the study, which revealed the following points:

- (a) $H_n(p)$ gave reasonable performance for $h_n \leq 0.50$;
- (b) When $0.05 < h_n < 0.50$, for each value of p , there is a range of widths h_n for which $H_n(p)$ has smaller mean square error. For large h_n values $h_n = 0.15$ gives the smallest discrepancy between the two estimators;
- (c) The two estimators $\hat{H}(p)$ and $H_n(p)$ do not perform well when p becomes large. The method of estimation has also been illustrated with a real data by Sankaran and Nair [515].

9.7 Estimation of Percentile Residual Life

As we have seen in the preceding chapters, the residual life distribution plays a fundamental role in inferring the lifetime remaining to a device given that it has survived a fixed time in operation. The percentiles of the residual life quantile function are the percentile residual life defined in (2.19) and its quantile form in (2.19). Classes of lifetime distributions based on monotone percentile residual life functions have been discussed in Sect. 4.3. The $(1-p)$ th percentile life function, according to the definition in (2.19), is

$$P(x) = Q[1 - p(1 - F(x))] - x, \quad x > 0. \quad (9.14)$$

As before, assume that $X_{1:n} \leq \dots \leq X_{n:n}$ are the ordered observations in a random sample of size n from the distribution with quantile function $Q(p)$. The sample analogue of (9.14) is then

$$p_n(x) = Q_n[1 - p(1 - F_n(x))] - x. \quad (9.15)$$

Csorgo and Csorgo [160] have discussed the asymptotic distribution of $p_n(x)$ for different cases consisting of $p_n(x)$, (a) as a stochastic process in x for fixed $0 < p < 1$, (b) as a stochastic process in p for a fixed $x > 0$, and (c) a two-parameter process in (p, x) . Using the density function of X , define

$$\begin{aligned} g(x) &= [f(P(x)) + x]^{-1}, \\ r_n(x) &= \frac{\sqrt{n}}{g(x)} [p_n(x) - P(x)], \\ H(x) &= B[1 - p(1 - F(x)) - pB(F(x))], \end{aligned}$$

where B is a Brownian bridge over $(0, 1)$. For a fixed $x > 0$, $H(x)$ is distributed as $N(0, p(1-p)(1-F(x)))$. For a fixed p , provided $g(p)$ is positive and continuous at $1 - p(1 - F(x))$, $r_n(x)$ is asymptotically $N(0, p(1-p)(1 - F(x)))$. In addition, if $f(x) > 0$ on $(Q(1-p), \infty)$ and some $r > 0$,

$$\sup_{x > Q(1-p)} \left[F(x)(1 - F(x)) \frac{|f'(x)|}{f^2(x)} \right] \leq \gamma$$

and $f(x)$ is ultimately non-increasing as $x \rightarrow \infty$, then almost surely

$$\sup_{0 < x < \infty} |r_n(x) - H(x)| = o(\delta_n)$$

with

$$\delta_n = \left(\frac{\log \log n}{n} \right)^{\frac{1}{4}} (\log n)^{\frac{1}{2}} \text{ as } n \rightarrow \infty.$$

A smooth version of the empirical estimator has been studied by Feng and Kulasekera [196]. Following this, as $n \rightarrow \infty$, Alam and Kulasekera [33] have established that under the above assumptions,

$$r_n(x) = \frac{1}{g(x)} \int_{-\infty}^{\infty} r_n(u) g(u) K\left(\frac{u-x}{\lambda}\right) \frac{du}{\lambda}$$

is asymptotically normal as

$$\bar{H}(x) = \frac{1}{g(x)} \int_{-\infty}^{\infty} H(u) g(u) K\left(\frac{u-x}{\lambda}\right) \frac{du}{\lambda}$$

and also

$$|\bar{r}_n(x) - \bar{H}(x)| = o(\delta_n)$$

almost surely. Since

$$E\bar{H}(x) = 0$$

and

$$\begin{aligned} E[\bar{H}(x)]^2 &= \frac{1}{g^2(x)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(u, w) g(u) g(w) K\left(\frac{u-x}{\lambda}\right) K\left(\frac{w-x}{\lambda}\right) \frac{du dw}{\lambda^2} \\ &= v_\lambda(x), \end{aligned} \tag{9.16}$$

the asymptotic normal distribution has mean zero and variance $v_\lambda(x)$. It is easy to see that the function $K(\cdot)$ denotes the kernel, which the authors assume to be a probability density function centred at the origin. The efficiency of the Csorgo's estimator $r_n(x)$ relative to that of $\bar{r}_n(x)$ is now

$$e = \frac{v_\lambda(x)}{p(1-p)(1-F(x))}.$$

The term $\mu(u, w)$ occurring in (9.16) is

$$\begin{aligned}\mu(u, w) &= EG(u)G(w) \\ &= p(1-p) + (p+p^2)(F(u) \wedge F(w)) \\ &\quad - p\{F(u) \wedge ((1-p)(1-F(w))) + F(w) \wedge ((1-p)(1-F(u)))\}.\end{aligned}$$

Assuming that

$$\int_{-\infty}^{\infty} g(u)K\left(\frac{p-x}{\lambda}\right)\frac{du}{\lambda} < \infty,$$

we have

$$\sup_{0 < x < \infty} E(r_n^2(x) - v^*(x)) = o(n^{-\frac{1}{4}})$$

and

$$\sup_{0 < x < \infty} E(\bar{r}_n^2(x) - v_\lambda^*(x)) = o(n^{-\frac{1}{4}})$$

as $n \rightarrow \infty$. In the above results,

$$v^*(x) = p[1 - p(1 - F(x))][1 - F(x)]$$

and

$$\begin{aligned}v_\lambda^*(x) &= \left(\frac{p}{\lambda g(x)}\right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(u)F(w)][1 - F(u)F(w)]g(u)g(w) \\ &\quad \times K\left(\frac{u-x}{\lambda}\right)K\left(\frac{w-x}{\lambda}\right)dudw + v_\lambda(x).\end{aligned}$$

The asymptotic value of the normalized difference between $\bar{r}_n(x)$ and the empirical estimator is given by

$$\frac{n[\text{MSE}(p_n(x)) - \text{MSE}(\bar{p}_n(x))]}{\lambda g^2(x)} = v_0^*(x),$$

where

$$\bar{p}_n(x) = \int_{-\infty}^{\infty} p_n(u)K\left(\frac{(u-t)}{\lambda}\right)\frac{du}{\lambda}.$$

Alam and Kulasekera [33] have also presented a Monte Carlo study when the underlying distribution is exponential and Weibull using uniform distribution over $[-1, 1]$ as the kernel. It has been observed through this study that the kernel estimator provides better results for moderate sample sizes and chosen values of λ .

More properties of $P_n(x)$ have been given by Csorgo and Mason [162], Aly [34] and Csorgo and Viharos [163].

Pereira et al. [491] have studied properties of the class of distributions with decreasing percentile residual life (DPRL). They introduced a nonparametric estimator of $P(x)$ based on the fact that

$$P(x) \text{ is DPRL} \Leftrightarrow P(x) = \inf_{y \leq x} P(y).$$

Thus, the estimator of $P(x)$ is given by

$$\tilde{P}(x) = I_{(x, \infty)}(X_{n:n}) \inf_{y \leq x} P_n(y), \tag{9.17}$$

where $I_{(x, \infty)}$ denotes the indicator function of the indicated interval. Note that $\tilde{P}(x)$ is the largest decreasing function that lies below the empirical $P_n(x)$. In practice, the estimator $\tilde{P}(x)$ can be computed easily in the following way. When $X_{1:n} \leq \dots \leq X_{n:n}$ are the ordered observations in a random sample of size n from the distribution $F(x)$, find the number of distinct values in the sample, say k . Let $Y_1 < \dots < Y_k$ be the resulting ordered values with no ties. Then, the estimate $\tilde{P}(x)$ is given by

$$\tilde{P}(x) = \begin{cases} P_n(Y_1-) + Y_1 - x & \text{if } x < Y_1 \\ \min\{P_n(Y_1-), P_n(Y_2-), P_n(Y_j-), P_n(Y_{j+1}-) + Y_{j+1} - x\} & \text{if } Y_j \leq x < Y_{j+1} \\ 0 & \text{if } x > Y_k. \end{cases}$$

The strong uniform consistency of the estimator $\tilde{P}(x)$ is presented in the following theorem.

Theorem 9.11. *Let X be a random variable having DPRL property. If the distribution function $F(x)$ of X has a continuous positive density function $f(x)$ such that $\inf_{0 \leq p \leq 1} f_x(Q(p)) > 0$, then $\tilde{P}(x)$ is a strongly uniformly consistent estimator of $P(x)$.*

Note that in order to estimate $p(x)$ under the condition that it increases, an estimator that is a modification of the estimator given in (9.15) can be obtained. It is also strongly uniformly consistent.

9.8 Modelling Failure Time Data

In this and in the subsequent sections, we consider various aspects of the process of modelling lifetime data using distributions. As a problem solving activity, the statistical concepts expressed in terms of quantile functions offer new perspectives

that are not generally available in the distribution function approach or at least provides an alternative approach with possibly different interpretations with almost equivalent results. Several factors have to be considered while constructing a model. Generally, the model builder will have some information about the phenomenon under consideration or will be able to extract some features from a preliminary assessment of the observations. The background information about the variables and possible distributions along with the necessary level of details required for the analysis are crucial points. Choice of the appropriate model also depends on the data available to ensure its adequacy and the method of estimation of the parameters. Finally, model parsimony is an attractive feature that prefers a simpler model to a more complex one. For example, models with lesser number of parameters or functional forms that have simpler structure (like constancy or linearity as against nonlinear) will be easier to build and analyse. Qualities such as tractability of the model, ease of analysis and interpretation are often prime considerations. This should be consistent with the ability of the model to represent the essential features of the life distribution that are inherent in the observations. In practice, there are three essential steps in building a model. They are identification of the appropriate model, fitting the model and finally checking its adequacy.

9.9 Model Identification

The procedure involved in model choice is to try out possible candidates and choose the best among them. We have seen in previous chapters (see Table 1.1 and the review of bathtub models in Chap. 7) that a plethora of lifetime distributions have been proposed to represent lifetimes in the distribution function approach. This adds to the complexity of determining the potential initial choice. A generalized version may fit in many practical situations, but more parsimonious solution that render easy analysis and interpretations may exist. The problem is somewhat of a lesser degree when quantile functions are used. We have the generalized lambda distribution (Sect. 3.2.1) or the generalized Tukey lambda family (Sect. 3.2.2) that can take care of a wide variety of practical problems in view of their ability to provide reasonable approximations to many continuous distributions. See the discussion on the structural properties of the two quantile functions. The reliability properties of the models, methods of estimating the parameters and examples of fitting them (Sect. 3.6) have been described in detail in Chap. 3. When there are multiple solutions that give models which fit the data, the one which captures the observed features of the reliability characteristics more closely may be preferred. The reliability characteristic may be the hazard quantile function, mean residual quantile function or any other for which the fitted model is put to good use. We can also make use of other models including those suggested in the distribution function approach with tractable quantile functions. A look at the admissible range of skewness and kurtosis values for the proposed distribution will indicate if it covers the distributional shape that fits the observations. The skewness and kurtosis

coefficients of the sample values have to be within the ranges prescribed for the chosen model.

Another useful method to arrive at a realistic model is to compare any of the basic reliability functions that uniquely determines the life distribution, with its sample counterpart. The hazard quantile function, mean residual quantile function, etc. can be used for this purpose. To use the hazard quantile function, recall its definition

$$H(u) = \frac{1}{(1-u)q(u)}.$$

Let $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be an ordered set of observations on failure times. Then, the quantile function of the distribution of $X_{r:n}$ is (1.26)

$$Q_r(u_r) \equiv Q(I^{-1}(r, n - r + 1)).$$

If U has a uniform distribution, then X , where $x = Q(p)$, has quantile function $Q(u)$. Hence, the ordered U_r , say $u_{(r)}$, leads to $x_{r:n} = Q(u_{(r)})$. So, as an approximation, either the mean

$$EQ(U_{(n)}) = Q\left(\frac{r-0.5}{n}\right),$$

or the median

$$M_r = Q(I^{-1}(0.5, r, n - r + 1)),$$

or equivalently

$$u_{(r)}^* = I^{-1}(0.5, r, n - r + 1), \quad u_{(r)}^* = Q^{-1}(M_r) = F(M_r),$$

can be used. Gilchrist [215] refers to the function I^{-1} as BETAINV and points out that it is a crucial standard function in most spreadsheets and statistical software. The empirical quantile density function $\hat{q}(u)$ can be obtained from the data, u_r , from the median probability. Thus, from the above formula, $q(u_{(r)}^*)$ can be plotted against $u_{(r)}^*$. Once a graph of

$$\hat{H}(u) = \frac{1}{(1-u_r^*)q(u_r^*)}$$

is obtained, its functional form can be obtained by comparing the plot with one of the hazard quantile function forms. Several such forms are available from Table 2.4, Chaps. 3 and 7.

A third alternative in model identification is to start with a simple model and then modify it to accommodate the features of the data. Various properties of quantile functions described in Sect. 1.2 can assist in this regard. For example, the

power distribution has an increasing $H(u)$, while the Pareto II has a decreasing hazard quantile function. The product of these two is the power-Pareto distribution discussed in Chap. 3, with a highly flexible form for $H(u)$. We can also use various kinds of transformations to arrive at new models from the initially assumed one. An excellent discussion of these methods along with various illustrations is available in Gilchrist [215].

9.10 Model Fitting and Validation

Once the data is collected and a specific model form is assumed, the next goal is to estimate the parameters. One of the methods of estimation discussed in the preceding sections of this chapter can be employed for this purpose. The only remaining step in the model building process is to ascertain whether the model with the estimated parameters describes the data adequately. This is called model validation. Since the parameters have been estimated from the data using some optimality criteria, the reproducibility of the model will be enhanced if its validation is made by another set of data if it exists. When the data is large, part of it can be used for identification and fitting while the remaining for validation. Sometimes, cross-validation is made use of wherein part of the data used for fitting and the remaining part of the data used for validation are interchanged and the two acts are repeated.

There are graphical methods to ascertain the goodness of fit. One is the $Q-Q$ plot and the other is the box plot mentioned in Chap. 1. The $Q-Q$ plots were illustrated in the modelling of real data using lambda distribution, the power-Pareto and the Govindarajulu distributions in Chap. 3; see, for example, Figs. 3.7–3.9. An advantage of the $Q-Q$ plot is that it can be used to specifically compare the tail areas. We can consider the plots comparing $x_{(r)}$ with its values at 0.90 or 0.95 or 0.05, using the median rankits

$$\hat{Q}(I^{-1}(0.05, r, n - r + 1)), \hat{Q}(I^{-1}(0.025, r, n - r + 1)), \hat{Q}(I^{-1}(0.975, r, n - r + 1)).$$

The recent work of Balakrishnan et al. [52] constructing optimal plotting points based on Pitman closeness and its performance as a good of fit and comparison with other plotting points is of special interest here. A second method is to apply some goodness-of-fit tests like chi-square. Suppose there are n observations and are divided into m groups each containing the same number of observations. Take $u_j = \frac{j}{m}$, $u_0 = 0$, $u_m = 1$, $r = 0, \dots, m - 1$. If $p_j = \hat{Q}(u_r)$ and f_j is the frequency of observations in (p_{r-1}, p_r) , the expected value of f_r is $\frac{n}{m}$ for all value of r . Then, the statistic

$$\sum \left[\frac{f_r - \left(\frac{n}{m}\right)^2}{\left(\frac{n}{m}\right)} \right]$$

has approximately a chi-square distribution with $n - 1$ degrees of freedom. This scheme is more easier to apply than the conventional chi-square procedure. For elaborate details on chi-squared tests and their power properties, one may refer to the recent book by Voinov et al. [575]. The general references on different forms of goodness-of-fit tests of D'Agostino and Stephens [165] and Huber-Carol et al. [289] will also provide valuable information in this regard. See also Gilchrist [214,215] for methods of estimators when quantile functions are used in modelling statistical data.