Chapter 5

INFERENCE FOR EXPONENTIATED SCALE AND LOCATION FAMILY OF DISTRIBUTIONS BASED ON GROUPED DATA

5.1 Introduction

Many times in a life testing problem, it is not possible to monitor units continuously; instead one inspects the units at pre-specified times. Thus, the data consists of the number of failures or deaths in an interval. For example, when testing a large number of inexpensive units for time to failure, it may be cost prohibitive to connect each one to monitoring device. Thus, an inspector may inspect them at predetermined time intervals and record the number of units that failed since the last inspection.

Similarly, in cancer follow up studies where the variable of interest is the time to relapse, a patient may be monitored only at regular intervals or may seek help only after tangible symptoms of the disease appear. Thus, the time to relapse cannot be specified exactly, but will only be known to lie between two successive clinic visits. In a number of industrial situations also, it is more economical to observe number of failures of components in predefined time intervals instead of observing exactly.

Group samples from exponential distribution were studied by various authors as well Nelson (1977), Chang and Chen (1988) and Shapiro and Gulati (1998).These authors studied the use of maximum likelihood methods to make inference about the parameters of the exponential and Weibull distributions.

In this chapter, we discuss the inference for ESL family of distributions based on grouped data. In Section 5.2, we provide the estimation of unknown parameters using maximum likelihood method. Confidence intervals and testing of base line distributions are discussed in Section 5.3. β-expectation tolerance interval and β-content γ-level tolerance interval for ES family of distributions are provided in Section 5.4. We apply the procedure to EE distribution considering third bus-motor failure data in Section 5.5.

In the following section, we obtain MLE based on grouped data for unknown parameters for ESL family of distributions.

5.2 Maximum Likelihood Estimator

When data from a continuous life distribution are grouped, estimation can be based on the exact likelihood function for the grouped data.

Suppose that there are n independent units under observation at time 0 and that lifetime Y_j of the j th unit has ESL family of distributions defined in (3.1.2), j=1,2,3...,n. If number of failures at time t_i , i=1,2,...,k are recorded then x_i , i=1,2,....,k, be the number of failures in the interval (t_{i-1}, t_i) with $t_0=0$ and $t_k=\infty$ such that $\sum_{i=1}^{k} x_i = n$. Let p_i be the probability of an observation falling into interval (t_{i-1}, t_i) and $\sum_{i=1}^{k} p_i = 1$. Thus, we have grouped data from the underlying distribution. The points t_i's, i=1,2,...,k are called group limits to be decided by experimenter.

The likelihood function for the unknown parameters is

$$L(\alpha,\mu,\theta;\underline{x}) = \frac{n!}{\prod_{i=1}^{k} x_i} \left(\prod_{i=1}^{k} (p_i(\alpha,\mu,\theta))^{x_i} \right),$$

where $p_i(\alpha,\mu,\theta) = F(t_i,\alpha,\mu,\theta) - \mathrm{F}(\mathrm{t}_{i-1},\alpha,\mu,\theta)$.

i,e. $p_i(\alpha,\mu,\theta) = (G(t_i))^{\alpha} - (G(t_{i-1}))^{\alpha}$, i=1,2,...,k,

where $G(t) = G\left(\frac{t-\mu}{\theta}\right)$, c.d.f. of the base line distribution with location parameter μ and scale parameter θ.

Estimates of the parameters of the ES family of distributions are obtained by maximizing the log-likelihood function. The log-likelihood equation of grouped data is given by $L = c + \sum_{i=1}^{k} x_i \ln\left((G(t_i))^{\alpha} - (G(t_{i-1}))^{\alpha}\right)$, (5.2.1)

where c is constant independent of α, μ and θ.

Therefore to obtain MLE of (α, μ, θ)', differentiating (5.2.1) partially with respect to α, μ and θ and equating it to zero. We have

$$\frac{\partial L}{\partial \alpha} = \sum_{1=1}^{k} \frac{x_i \ln G(t_i)}{1 - \left(\frac{G(t_{i-1})}{G(t_i)}\right)^{\alpha}} - \sum_{i=1}^{k} \frac{x_i \ln G(t_{i-1})}{\left(\frac{G(t_i)}{G(t_{i-1})}\right)^{\alpha} - 1} = 0,$$

$$\frac{\partial L}{\partial \mu} = \sum_{1=1}^{k} \frac{\alpha\, x_i \ln G'_{\mu}(t_i)}{G(t_i)\left(1 - \left(\frac{G(t_{i-1})}{G(t_i)}\right)^{\alpha}\right)} - \sum_{1=1}^{k} \frac{\alpha\, x_i \ln G'_{\mu}(t_{i-1})}{G(t_i)\left(\left(\frac{G(t_i)}{G(t_{i-1})}\right)^{\alpha} - 1\right)} = 0 \text{ and}$$

$$\frac{\partial L}{\partial \theta} = \sum_{1=1}^{k} \frac{\alpha\, x_i \ln G'_{\theta}(t_i)}{G(t_i)\left(1 - \left(\frac{G(t_{i-1})}{G(t_i)}\right)^{\alpha}\right)} - \sum_{1=1}^{k} \frac{\alpha\, x_i \ln G'_{\theta}(t_{i-1})}{G(t_i)\left(\left(\frac{G(t_i)}{G(t_{i-1})}\right)^{\alpha} - 1\right)} = 0.$$

It is clear that no closed form solution is possible. We can use Newton-Raphson method to solve the above three equations. Asymptotic confidence interval and testing on shape parameter are discussed in the following section.

5.3 Confidence Interval and Testing

It is important to note, however, that the asymptotic theory of the likelihood method is for the solutions of the likelihood equations. This includes the asymptotic normality. We now state the theorem, which establishes the asymptotic normality of the estimators.

Theorem 5.1: As $n \to \infty$, $\sqrt{n}\left((\hat{\alpha}-\alpha),(\hat{\mu}-\mu),(\hat{\theta}-\theta)\right) \to N_3(0, I^{-1}(\alpha,\mu,\theta))$, where $I^{-1}(\alpha,\mu,\theta)$ is the inverse of Fisher information matrix $I(\alpha,\mu,\theta)$, of the so-obtained MLE $(\hat{\alpha}, \hat{\mu}, \hat{\theta})'$.

$$I(\alpha,\mu,\theta) = \begin{bmatrix} E\left(\frac{\partial^2 L}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 L}{\partial \alpha \partial \mu}\right) & E\left(\frac{\partial^2 L}{\partial \alpha \partial \theta}\right) \\ E\left(\frac{\partial^2 L}{\partial \mu \partial \alpha}\right) & E\left(\frac{\partial^2 L}{\partial \mu^2}\right) & E\left(\frac{\partial^2 L}{\partial \mu \partial \theta}\right) \\ E\left(\frac{\partial^2 L}{\partial \theta \partial \alpha}\right) & E\left(\frac{\partial^2 L}{\partial \theta \partial \mu}\right) & E\left(\frac{\partial^2 L}{\partial \theta^2}\right) \end{bmatrix}.$$

It is not possible to obtain the Fisher information matrix in a closed form. $I^{-1}(\alpha,\mu,\theta)$ will be estimated by $I^{-1}(\hat{\alpha}, \hat{\mu}, \hat{\theta})$ and this can be used to construct the asymptotic 100(1-δ)% confidence interval of α, μ and θ are respectively given by $\hat{\alpha} \pm z_{1-\delta/2}\sqrt{\frac{I_{\alpha\alpha}^{-1}}{n}}$, $\hat{\mu} \pm z_{1-\delta/2}\sqrt{\frac{I_{\mu\mu}^{-1}}{n}}$ and $\hat{\theta} \pm z_{1-\delta/2}\sqrt{\frac{I_{\theta\theta}^{-1}}{n}}$, where $z_{1-\delta/2}$ is the $(1-\delta/2)^{th}$ quantile of the standard normal distribution and $I^{-1}_{\alpha\alpha}, I^{-1}_{\mu\mu}$ and $I^{-1}_{\theta\theta}$ are the diagonal elements of matrix $I^{-1}(\hat{\alpha}, \hat{\mu}, \hat{\theta})$.

The asymptotic normality is also useful for testing the goodness of fit of the model using likelihood ratio test based on MLE. We shall describe briefly as follows.

Consider the above grouped data setup and consider the hypothesis:

$H_0 : p_i = p_{i0}$ i=1,2,...,k, where the p_{i0}'s are specified but may involve unknown parameters. Let $\tilde{p}_{i0}$ be the MLE of p_i under H_0.

The likelihood ratio statistic for testing H_0 against the alternative that the p_i's satisfy only $p_i \geq 0, \sum p_i = 1$, is easily seen to be

$$\Lambda = 2\sum_{i=1}^{k} x_i \ln\left(\frac{x_i}{n}\right) - 2\sum_{i=1}^{k} x_i \ln \tilde{p}_{i0} .$$

The limiting distribution of Λ under H_0 is $\chi^2_{(k-s)}$ when the p_{i0}'s involves unknown parameters.

By using MLE $(\hat{\alpha}, \hat{\mu}, \hat{\theta})'$ and similarly obtained MLE $(\tilde{\mu}, \tilde{\theta})'$ by modifying likelihood by taking α=1, one can construct the likelihood ratio goodness-of-fit test of baseline model.

One sided tolerance intervals when f(x,α,θ) is a two parameter ES family of distributions, based on maximum likelihood estimators for ungrouped data were discussed by Shirke et al. (2005). In the following, we extend the procedure for setting tolerance interval based on grouped data.

5.4 Tolerance Intervals

In this subsection, we obtain tolerance interval namely β-expectation tolerance interval and β-content γ-level tolerance interval based on grouped data for ESL family of distributions with known location parameter μ. Without loss of generality, consider $\mu=0$ in (3.1.2) model.

The β^{th} percentile of (3.1.2) is denoted by $X_\beta(\alpha,\theta)$ and is given by $X_\beta(\alpha,\mu,\theta)=\theta\, G^{-1}(\beta^{1/\alpha})$, where $G^{-1}(.)$ is the inverse function of $G(.)$. Since the parameter $(\alpha,\theta)'$ is unknown, by invariance property of MLE, we replace it by its MLE $(\hat{\alpha},\hat{\theta})'$. Hence an approximate upper β-expectation tolerance interval $I_1(\underline{X})$ is $I_1(x)=(0,\hat{\theta}\, G^{-1}(\beta^{1/\hat{\alpha}}))$. (5.4.1)

Expected coverage of $I_1(\underline{X})$ is given by following theorem due to Shirke et al. (2005).

Theorem 5.1: An expected coverage of $I_1(\underline{X})$ is given by

$$E[F\{X_\beta(\hat{\alpha},\theta);\alpha,\theta)] = \beta + A(\alpha,\theta)\sigma_1^2 + B(\alpha,\theta)\sigma_2^2 + C(\alpha,\theta)\sigma_{12}. \quad (5.4.2)$$

where $F(\cdot)$ is defined in (3.1.1),

$$A(\alpha,\theta)=0.5\alpha(1-\alpha)[G(t)]^{\alpha-2}[G'_\theta(t)]^2+\alpha[G(t)]^{\alpha-1}\left[\frac{G''_{x\theta}(t)G'_\theta(t)}{G'_x(t)}-0.5G''_{\theta\theta}(t)\right],$$

$$B(\alpha,\theta)=\alpha^{-1}[G(t)]^\alpha\log G(t)\,[1+0.5\alpha\log G(t)],$$

$$C(\alpha,\theta)=[G(t)]^\alpha\log G(t)\left[\frac{(\alpha-1)G'_\theta(t)}{G(t)}+\frac{G''_{x\theta}(t)}{G'_x(t)}\right],$$

$$G(t)=G\left(\frac{t}{\theta}\right),\ G'_x(t)=\frac{\partial G(t)}{\partial x},\ G'_\lambda(t)=\frac{\partial G(t)}{\partial\lambda},\ G''_{x\theta}(t)=\frac{\partial^2G(t)}{\partial x\partial\theta},$$

$G''_{\theta\theta}(t)=\dfrac{\partial^2G(t)}{\partial\theta^2}$ while σ_1^2, σ_2^2 are asymptotic variances of $\hat{\theta}$ and $\hat{\alpha}$ respectively and $\sigma_{12}=\mathrm{cov}(\hat{\theta},\hat{\alpha})$ is an asymptotic covariance of $\hat{\theta}$ and $\hat{\alpha}$.

Proof : Follows from the result of Atwood (1984). We omit the same for brevity.

In the following, we obtain β-content γ-level tolerance interval.

Let $I_2(\underline{X})=(0,\ \delta\hat{X}_\beta)$ be an upper β-content γ-level tolerance interval for the distribution having c.d.f. (3.1.1). The factor $\delta>0$ is to be determined such that $I_2(\underline{X})$ is a β-content γ-level tolerance interval for $\beta\in(0,1)$ and $\gamma\in(0,1)$. i.e. $P\left[F\left(\delta\hat{X}_\beta;\theta,\alpha\right)\geq\beta\right]=\gamma$.

This implies $$P\left[\hat{X}_\beta \geq \frac{\theta G^{-1}\left(\beta^{1/\alpha}\right)}{\delta}\right] = \gamma. \tag{5.4.2}$$

Assuming consistency and asymptotic normality of the MLE $(\hat{\alpha},\hat{\theta})'$, we have $\hat{X}_\beta(\alpha,\theta)$ is asymptotically normal with mean $X_\beta(\alpha,\theta)$ and variance $\frac{\sigma^2(\alpha,\theta)}{n}$, where $\sigma^2(\alpha,\theta) = H\Sigma H'$ with Σ as a variance covariance matrix of $(\hat{\alpha},\hat{\theta},)'$ and $H = \left[\frac{\partial x_\beta(\alpha,\theta)}{\partial\alpha}, \quad \frac{\partial x_\beta(\alpha,\theta)}{\partial\theta}\right]$.

If Z is a standard normal variate then we can write from (5.4.2)

$$P\left[Z \leq \left\{\frac{\theta G^{-1}(\beta^{1/\alpha})}{\delta} - x_\beta(\alpha,\theta)\right\}\left(\frac{\sqrt{n}}{\sigma(\alpha,\theta)}\right)\right] = 1-\gamma.$$

Suppose $z_{1-\gamma} = \left\{\frac{\theta G^{-1}(\beta^{1/\alpha})}{\delta} - x_\beta(\alpha,\theta)\right\}\left(\frac{\sqrt{n}}{\sigma(\alpha,\theta)}\right)$, where $z_{1-\gamma}$ is the $100(1-\gamma)^{th}$ percentile of the standard normal variate then we have

$$\delta = \left[1 - \frac{z_{1-\gamma}\sigma(\alpha,\theta)}{\sqrt{n}\,\theta\, G^{-1}(\beta^{1/\alpha})}\right]^{-1}.$$

Note that δ depends on both the parameters θ and α. Replacing θ and α in δ by their respective MLE an asymptotic upper β-content γ-level tolerance interval $I_2(\underline{X})$ is $\left(0,\left[1+\frac{z_{1-\gamma}\sigma(\hat{\alpha},\hat{\theta})}{\sqrt{n}\ \hat{\theta}\,G^{-1}\left(\beta^{1/\hat{\alpha}}\right)}\right]^{-1}\hat{\theta}\,G^{-1}\left(\beta^{1/\hat{\alpha}}\right)\right)$. (5.4.3)

5.5 Application

Gupta and Kundu (2001a) have defined EE distribution in the following way. Let Y be a two parameter EE random variable with distribution function $F(y;\alpha,\lambda)=(1-e^{-\lambda y})^{\alpha}$ $\alpha>0$, $\lambda>0$, $y>0$. (5.5.1)

Therefore, the log-likelihood function for grouped data set up can be written as $L(\alpha,\lambda/\underline{x})=C_1+\sum_{i=1}^{k}x_i \ln\left[(1-e^{-\lambda t_i})^{\alpha}-(1-e^{-\lambda t_{i-1}})^{\alpha}\right]$, (5.4.2)

where C_1 is independent of α and λ. The MLE $\hat{\alpha}$ and $\hat{\lambda}$ can be obtained by maximizing equation (5.4.2) numerically with respect to λ and α.

Illustration 1: The analysis of bus-motor failure data was first considered by Davis (1952) and then revisited by Mudholkar et al. (1995). The data describe the number of miles (in thousands) to the first (of 191 buses) and the four succeeding major motor failures after appropriate repairs. The term failure refers to either abrupt breakdown or an event in which the maximum power produced by the bus fell below a fixed percentage of the normal rated value. The recorded data are the number of failures in interval of length

20,000 miles and reproduced in Table 5.1. Davis (1952) showed that the exponential distribution provides reasonable fits to the third failure data.

Table 5.1: Third bus-motor failure data.

Class Interval (1,000 miles)	Observed Frequency
0- 20	27
20- 40	16
40- 60	18
60-80	13
80- 100	11
100- up	16

We analyze third bus-motor failure data using exponential, gamma, Weibull and EE distributions. We estimate values of MLE of unknown parameters and log-likelihood (LL) for these distributions. The results are stated below.

For exponential distribution, MLE $\hat{\lambda}$=0.0167 and LL= -178.8222.

For gamma distribution, MLE $\hat{\alpha}$=1.1971, $\hat{\lambda}$ =49.2753 and LL= -178.3618.

For Weibull distribution, MLE $\hat{\alpha}$ =0.0091, $\hat{\lambda}$=1.1426 and LL= -178.1572.

For EE distribution, MLE $\hat{\alpha}$= 1.1971, $\hat{\lambda}$=0.0189 and LL= -178.4079.

It is clear that, we can use EE distribution as an alternative to gamma and Weibull distributions for this grouped data.

For EE distribution, we have computed the 95% asymptotic confidence interval for α as (1.1495, 1.2447) and that of λ as (0.0183, 0.0195). Also using percentile bootstrap method 95% asymptotic confidence interval for α as (0.8079, 1.7610) and that of λ as (0.0130, 0.0259).

The approximate upper β-expectation tolerance interval for EE distribution is given by $I_1(\underline{x}) = \left(0, \hat{\lambda}^{-1} \ln(1-\beta^{1/\hat{\alpha}})\right)$. Hence for above data, upper β-expectation tolerance limits (say $U_1(\underline{X})$) for various values of β's are

β	: 0.90	0.95	0.975	0.99
$U_1(\underline{X})$	: 112.6540	146.0220	179.3624	223.4168

Upper β-content γ-level tolerance interval given in equation (5.4.3) requires asymptotic variance covariance matrix of the MLE's $\hat{\alpha}$ and $\hat{\lambda}$ which involve complicated integrals to be solved numerically. This problem is resolved by using bootstrap technique.

According to bootstrap technique, we generate 5000 random samples each of size 101 from EE($\hat{\alpha}, \hat{\lambda}$) where $\hat{\alpha}$ and $\hat{\lambda}$ are estimated MLE for bus-motor failure data. For each of the bootstrap samples after grouping we have computed MLE of α and λ. Based on such 5000 MLEs, we obtain variance covariance matrix of MLE which is used to propose asymptotic upper β -content γ level tolerance limit (TL).and tabulated in Table 5.2.

Table 5.2: An asymptotic upper β -Content γ-level TL

β	$\sigma(\alpha,\lambda)$	γ=.90		γ=.95	
		δ	U($\underline{X}$)	δ	U($\underline{X}$)
0.90	15.5214	0.9985	130.7000	0.9981	130.6450
0.95	21.3797	0.9984	167.5300	0.9979	167.4542
0.975	27.4383	0.9983	204.2394	0.9978	204.1420
0.99	35.5961	0.9982	252.6830	0.9977	252.5567

5.6 Suggestions for Further Study

We propose to do the following in future.

- Bayesian and non-Bayesian estimation for exponentiated family of distributions based on record values.
- Test for the mixing proportion in the mixture of a degenerate and exponentiated family of distributions. UMVUE and Bayes estimate of reliability of mixed failure time distribution.
- Bivariate exponentiated family of distributions.
- Effect of mis-specification of the model on the inference and other statistical procedures like prediction.