The (P-A-L) Generalized Exponential Distribution: Properties and Estimation

M.R. Mahmoud and R.M. Mandouh

Institute of Statistical Studies & Research, Cairo University, Egypt

Abstract

In this paper, The PAL generalized exponential distribution is introduced as a new lifetime distribution. Some properties of the new distribution are studied. The maximum likelihood estimates and asymptotic variance-covariance matrix are obtained. Also, approximate Bayes estimates are computed using the Gibbs sampling procedure. Finally, an application to real data set is given.

Keywords: The PAL family; maximum likelihood estimates; asymptotic variance-covariance matrix; TTT plot; Kaplan-Meier estimate; Bayesian estimation

1. Introduction

Adding parameters to an existing distribution to generate a “generalized” distribution is a very common approach for developing more flexible distributions. There are many of generalized families that can be obtained by adding one parameter such as exponentiated family (Mudholkar and Sirvastava, 1993), Marshall and Olkin family (Marshall and Olkin, 1997), transmuted family (Shaw & Buckley, 2009), The Kw-G family which can be defined as an exponentiated family (Cordeiro and de Castro, 2011), The PAL family (Pappas et al., 2012) and others. Here we will be concerned with the last one.

2. The (P-A-L) Generalized Exponential Distribution

Pappas et al. (2012) introduced a new family to generalize a distribution by adding a shape parameter, namely PAL family. This family takes the following form
\[ s(x) = \frac{\ln\{1 - (1 - p)s_0(x)\}}{\ln(p)}, \quad x \in \mathbb{R}, \quad p \in \mathbb{R}^+ - \{0\} \quad (1) \]

and when \( p \to 1 \), then \( s \to s_0 \). The pdf and hazard function will be

\[ f(x) = \frac{(p - 1)f_0(x)}{\{1 - (1 - p)s_0(x)\}\ln(p)}, \quad (2) \]

\[ h(x) = \frac{(p - 1)f_0(x)h_0(x)}{\{1 - (1 - p)s_0(x)\}[\ln[1 - (1 - p)s_0(x)]]}, \quad (3) \]

where \( f_0 \) and \( h_0 \) are the pdf and hazard function of the base distribution. Pappas et al. (2012) introduced the (P-A-L) extended modified Weibull distribution. Al-Zahrani et al. (2015) studied the (P-A-L) extended Weibull distribution.

Now, we will take \( s_0 = [1 - (1 - e^{-\lambda x})^\alpha] \) which is the survival function of generalized exponential distribution introduced by Gupta and Kundu (1999).

So, inserting this into (1) gives

\[ s(x) = \frac{\ln\{1 - (1 - p)[1 - (1 - e^{-\lambda x})^\alpha]\}}{\ln(p)}, \quad (4) \]

Hence

\[ f(x) = \frac{\alpha \lambda (p - 1)(1 - e^{-\lambda x})^{\alpha - 1}e^{-\lambda x}}{\{1 - (1 - p)[1 - (1 - e^{-\lambda x})^\alpha]\}\ln(p)^{-1}}, \quad x, \lambda, \alpha > 0, \quad p \in \mathbb{R}^+ - \{0, 1\} \quad (5) \]

The hazard function is given by

\[ h(x) = \frac{\alpha \lambda (p - 1)(1 - e^{-\lambda x})^{\alpha - 1}e^{-\lambda x}}{\{1 - (1 - p)[1 - (1 - e^{-\lambda x})^\alpha]\}\ln[1 - (1 - p)[1 - (1 - e^{-\lambda x})^\alpha]]}, \quad (6) \]

Figures 1-2 illustrate some of the possible shapes of the pdf and hazard function of the PAL generalized exponential distribution for selected values of the parameters \( \alpha, \lambda \) and \( p \).
Now, to find the raw moments, we have

\[ E(X^r) = \alpha \lambda (p - 1) \int_0^\infty \frac{x^r (1 - e^{-\lambda x})^{\alpha - 1} e^{-\lambda x}}{\{(1 - (1 - p)(1 - e^{-\lambda x})^\alpha)\ln(p)\}} \, dx \]

Numerical integration procedures can be used to calculate the \( r \)th raw moments of the PAL generalized exponential distribution. Also, one can use Binomial expansion to obtain the following expression:

\[ E(X^r) = \frac{\alpha (p-1)^r (r+1)}{\ln(p) \lambda^r} \sum_{j=0}^\infty \sum_{k=0}^j \sum_{i=0}^\infty (-1)^{k+i} \binom{j}{k} \binom{\alpha (k+1)}{i} (1-p)^j \frac{1}{(i+1)^{r+1}}. \]
3. Random Number Generation and Parameter Estimation

Using inversion method, one can generate random from the PAL generalized exponential with the following formula

\[ x = \frac{-1}{\lambda} \ln \left( 1 - \frac{\alpha e^{\ln p} - p}{1 - p} \right), \]

where \( u \) distributed as uniform distribution. Now, Parameter estimation using maximum likelihood and Bayesian method will be discussed.

3.1 Maximum Likelihood Estimation

Let \( x_1, x_2, x_3, \ldots, x_n \) be a random sample follow the P-A-L Generalized Exponential distribution. The likelihood function is given by

\[ L(x; \alpha, \lambda, p) = \prod_{i=1}^{n} \frac{\alpha \lambda (p - 1) e^{-\lambda x_i} (1 - e^{-\lambda x_i})^{\alpha - 1}}{\ln(p) [1 - (1 - p) [1 - (1 - e^{-\lambda x_i})^\alpha]]} \]

and the log likelihood takes the form

\[ \ln L = n \ln(\alpha) + n \ln(\lambda) + n \ln(p - 1) - n \ln(\ln(p)) - \sum_{i=1}^{n} \lambda x_i \]

\[ - (\alpha - 1) \sum_{i=1}^{n} \ln(1 - e^{-\lambda x_i}) \]

\[ - \sum_{i=1}^{n} \ln([1 - (1 - p)(1 - (1 - e^{-\lambda x_i})^\alpha)]) \]  

Differentiating (8) with respect to \( \alpha, \lambda \) and \( p \), we have

\[ \frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln(1 - e^{-\lambda x_i}) \]

\[ - \sum_{i=1}^{n} \frac{(1 - p)(1 - e^{-\lambda x_i})^\alpha \ln(1 - e^{-\lambda x_i})}{[1 - (1 - p)[1 - (1 - e^{-\lambda x_i})^\alpha]]}, \]

\[ \frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} + (\alpha - 1) \sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i}}{(1 - e^{-\lambda x_i})} \]

\[ - \alpha (1 - p) \sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i} (1 - e^{-\lambda x_i})^{\alpha - 1}}{[1 - (1 - p)[1 - (1 - e^{-\lambda x_i})^\alpha]]} \]
and
\[
\frac{\partial \ln L}{\partial p} = \frac{n}{p - 1} - \frac{n}{p \ln(p)} \sum_{i=1}^{n} \frac{[1 - (1 - e^{-\lambda x_i})^\alpha]}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]}.
\]

Equating the derivatives in (9), (10) and (11) to zero and solve the three nonlinear equations numerically, we obtain the maximum likelihood estimators \(\hat{\alpha}, \hat{\lambda}, \hat{p}\).

The second derivatives of the logarithms of likelihood function is given by
\[
\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{n}{\alpha^2} - p(1 - p) \sum_{i=1}^{n} \frac{(1 - e^{-\lambda x_i})\ln(1 - e^{-\lambda x_i})}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2},
\]
\[
\frac{\partial^2 \ln L}{\partial \alpha \partial \lambda} = \sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i}}{(1 - e^{-\lambda x_i})} - (1 - p) \sum_{i=1}^{n} \frac{e^{-\lambda x_i}(1 - e^{-\lambda x_i})^{\alpha - 1}}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2}
\]
\[
+ \alpha(1 - p)^2 \sum_{i=1}^{n} \frac{x_i e^{-\lambda x_i}(1 - e^{-\lambda x_i})^{2\alpha - 1}}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2},
\]
\[
\frac{\partial^2 \ln L}{\partial p \partial \alpha} = \sum_{i=1}^{n} \frac{(1 - e^{-\lambda x_i})\ln(1 - e^{-\lambda x_i})}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2}
\]
\[
\frac{\partial^2 \ln L}{\partial p \partial \lambda} = \sum_{i=1}^{n} \frac{\alpha x_i e^{-\lambda x_i}(1 - e^{-\lambda x_i})^{\alpha - 1}}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2},
\]
\[
\frac{\partial^2 \ln L}{\partial p^2} = -\frac{n}{(p - 1)^2} + \frac{2n}{p^2 \ln(p)^2} + \sum_{i=1}^{n} \frac{[1 - (1 - e^{-\lambda x_i})^\alpha]^2}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2},
\]
and
\[
\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n}{\lambda^2} - (\alpha - 1) \sum_{i=1}^{n} \frac{x_i^2 e^{-\lambda x_i}}{(1 - e^{-\lambda x_i})^2}
\]
\[
- \alpha(1 - p) \sum_{i=1}^{n} \frac{x_i^2 e^{-\lambda x_i}(1 - e^{-\lambda x_i})^{\alpha - 2}}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2}
\]
\[
+ \alpha^2(1 - p)^2 \sum_{i=1}^{n} \frac{x_i^2 (e^{-\lambda x_i})^2(1 - e^{-\lambda x_i})^{2\alpha - 2}}{[1 - (1 - p)(1 - e^{-\lambda x_i})^\alpha]^2}.
\]

To obtain interval estimation of the parameters\((\alpha, \lambda, p)\), we first obtain the
3×3 observed information matrix \(J(\theta)\) which take the form:
\( J(\theta) = - \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\lambda} & I_{\alpha p} \\ I_{\alpha\lambda} & I_{\lambda\lambda} & I_{\lambda p} \\ I_{\alpha p} & I_{\lambda p} & I_{pp} \end{pmatrix} \)

whose elements are the second derivatives of the logarithms of likelihood function which are given above. An \( 100(1-\gamma)\% \) approximate confidence interval for each parameter \( \theta_i \) is given by

\[
(\hat{\theta}_i - Z_{\gamma/2} \sqrt{\hat{J}_{\theta_i\theta_i}}, \hat{\theta}_i + Z_{\gamma/2} \sqrt{\hat{J}_{\theta_i\theta_i}}),
\]

where \( \hat{J}_{\theta_i\theta_i} \) is the (i, i) diagonal element of \( \hat{J}(\hat{\theta})^{-1} \) (approximate variance-covariance matrix) for \( i = 1, 2, 3 \) and \( Z_{\gamma/2} \) is the quantile \( 1 - \gamma/2 \) of the standard normal distribution.

### Application to Real Data

**Data Set:** The following data is an uncensored data set consisting of 100 observations on breaking stress of carbon fibers (in Gba): 0.92, 0.928, 0.997, 0.9971, 1.061, 1.117, 1.162, 1.183, 1.187, 1.192, 1.196, 1.213, 1.215, 1.2199, 1.22, 1.224, 1.225, 1.228, 1.237, 1.24, 1.244, 1.259, 1.261, 1.263, 1.276, 1.31, 1.321, 1.329, 1.331, 1.337, 1.351, 1.359, 1.388, 1.408, 1.449, 1.4497, 1.45, 1.459, 1.471, 1.475, 1.477, 1.48, 1.489, 1.501, 1.507, 1.515, 1.53, 1.5304, 1.533, 1.544, 1.5443, 1.552, 1.556, 1.562, 1.565, 1.585, 1.586, 1.599, 1.602, 1.614, 1.616, 1.617, 1.628, 1.684, 1.711, 1.718, 1.733, 1.738, 1.743, 1.759, 1.777, 1.794, 1.799, 1.806, 1.814, 1.816, 1.828, 1.83, 1.884, 1.892, 1.944, 1.972, 1.984, 1.987, 2.02, 2.0304, 2.029, 2.035, 2.037, 2.043, 2.046, 2.059, 2.111, 2.165, 2.686, 2.778, 2.972, 3.504, 3.863, 5.306.

The maximum likelihood estimates for the P A L generalized exponential distribution are given by

\[ \hat{\alpha} = 48.4364, \quad \hat{\lambda} = 2.12578, \quad \hat{p} = 0.033487 \]

Through The popular Kolmogorov-Smirnov goodness of fit test, we fit the P A L generalized exponential distribution to this data and we have Kolmogorov-Smirnov test statistic \( D=0.0915 \) with \( p \)-value \( 0.373 > 0.05 \). So, we had no reason to reject the null hypothesis that generated data follows the P A L generalized exponential distribution.

In our example, the 95\% confidence intervals for the three parameters \( \alpha, \lambda, p \) are listed in the following table.

**Table 1:** MLEs and confidence intervals of parameters in the case of the PAL generalized exponential model based on 100 breaking stress data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>48.4364</td>
<td>12.583</td>
<td>(23.774, 73.099)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>2.12578</td>
<td>0.2983</td>
<td>(1.5411, 2.7104)</td>
</tr>
<tr>
<td>( p )</td>
<td>0.033487</td>
<td>0.0003</td>
<td>(0.0735, 0.1405)</td>
</tr>
</tbody>
</table>
To obtain the MLE of the survival function $S(x)$ of the PAL generalized exponential distribution, replace the parameters $a$, $\lambda$ and $p$ by their MLEs $\hat{a}$, $\hat{\lambda}$, and $\hat{p}$ in (4) as follow:

$$\hat{S}(x) = \frac{\ln\{1 - (1 - \hat{p}) \left[ 1 - \left( 1 - e^{-\hat{\lambda}x} \right)^{\hat{a}} \right]\}}{\ln(\hat{p})}$$

Kaplan and Meier (1958) suggested the following estimation procedure of reliability function (survival function) for a nonparametric model: Fix $t > 0$. Let $t_{(1)} < t_{(2)} < \cdots < t_{(n)}$ denote the recorded functioning times, either until failure or to censoring, ordered according to size. Let $J_t$ denote the set of all indices $j$ where $t_{(j)} \leq t$ represents a failure time. Let $n_j$ denote the number of units, functioning and in observation immediately before time $t_{(j)}$, $j = 1, 2, \ldots, n$. Then the Kaplan-Meier estimator (KME) of $S(t)$ is defined as:

$$\hat{S}(t) = \prod_{j \in J_t} \frac{n_j - 1}{n_j}$$

For more details about properties of Kaplan-Meier estimator see Høyland and Raussand (1994).

Now, the MLE and Kaplan-Meier estimator of reliability function are obtained for our data set. The estimates are displayed graphically in the following figures:

![Figure 1](image)

Figure 1: $R$ is the MLE of reliability function for complete data, $S$ is KME of reliability function for complete data

The empirical scaled TTT transform (Aarset (1987)) can be used to identify the shape of the hazard function. The scaled TTT transform is convex (concave) if the hazard rate is decreasing (increasing), and for bathtub (unimodal) hazard rates, the scaled TTT transform is first convex (concave) and then concave (convex).
The TTT plot for complete data is the plot of \( \left( \frac{i}{n}, G \left( \frac{i}{n} \right) \right) \), where \( G \left( \frac{i}{n} \right) = \frac{\sum_{j=1}^{i} T_{j,n} + (n-i)T_{i,n}}{\sum_{j=1}^{n} T_{j,n}} \) for \( i = 1, 2, ..., n \). \( \sum_{j=1}^{n} T_{j,n} \) is the total time on test at the \( i \)th failure for \( i = 1, 2, ..., n \) and \( T_{j,n}, j = 1, 2, ..., n \), are the order statistics of the sample. Figure 1: present TTT of complete data.

As displayed in figure 1: the TTT plot has increasing shaped failure rate which agrees with the plot of the MLE of failure rate.

### 3.2 Bayesian Estimation

In this section, approximate Bayes estimates are computed using the Gibbs sampling procedure. This procedure is used to generate samples from the posterior distributions. The approximate Bayes estimators are obtained under the assumptions of non-informative priors.

We consider the PAL generalized exponential model with density function (5) and a non-informative joint prior distribution for \( \alpha, \lambda \) and \( p \) given by:

\[
\pi_0(\alpha, \lambda, p) \propto \frac{1}{a\lambda p},
\]

where \( \alpha, \lambda \) and \( p > 0 \). The joint posterior distribution for these parameters can be written as

\[
\pi(\alpha, \lambda, p | x) \propto \pi_0(\alpha, \lambda, p)\exp\{l(x; \alpha, \lambda, p)\}
\]

where \( l(x; \alpha, \lambda, p) \) is the logarithm of the log likelihood function given by (8), which is

\[
l(x; \alpha, \lambda, p) = n\ln(\alpha) + n\ln(\lambda) + n\ln(p - 1) - n\ln(\ln(p)) - \sum_{i=1}^{n} \lambda x_i - (\alpha - 1) \sum_{i=1}^{n} \ln(1 - e^{-\lambda x_i}) - \sum_{i=1}^{n} \ln([1 - (1 - p)][1 - (1 - e^{-\lambda x_i})^p])
\]

The reparametrization \( \rho_1 = \log(\alpha), \rho_2 = \log(\lambda), \) and \( \rho_3 = \log(p) \) are considered. One can obtain from (12) a non-informative prior for \( \rho_1, \rho_2, \) and \( \rho_3 \) as follow

\[
\pi(\rho_1, \rho_2, \rho_3) = \text{constant}, \quad \text{where} \quad -\infty < \rho_1, \rho_2 \text{and} \rho_3 < \infty.
\]
The convergence of the Gibbs sampling algorithm is obtained by the choice of the values of hyper-parameters of the uniform priors.

Using the above reparametrization, the joint posterior distributions for $\rho_1$, $\rho_2$ and $\rho_3$ is

$$
\pi(\rho_1, \rho_2, \rho_3 \mid x) \propto \pi(\rho_1, \rho_2, \rho_3). \exp \left\{ n\rho_1 + n\rho_2 + n.\ln[\exp(\rho_3) - 1] - n.\ln\rho_3 - \sum_{i=1}^{n} \exp(\rho_2) \cdot x_i - (\exp(\rho_3) - 1) \sum_{i=1}^{n} \ln[1 - \exp(-\exp(\rho_2) x_i)] - \sum_{i=1}^{n} \ln[1 - (1 - \exp(\rho_3))(1 - (1 - e^{-\exp(\rho_2) x_i})^{\exp(\rho_1)})] \right\} \tag{14}
$$

Assuming the prior $\pi(\rho_1, \rho_2, \rho_3) = \text{costant}$, the conditional posterior distributions used in the Gibbs sampling algorithm are given by:

$$
\pi(\rho_1 \mid \rho_2, \rho_3, x) \propto \exp \left\{ n\rho_1 + n.\ln[\exp(\rho_3) - 1] - n.\ln\rho_3 - \sum_{i=1}^{n} \exp(\rho_2) \cdot x_i - (\exp(\rho_3) - 1) \sum_{i=1}^{n} \ln[1 - \exp(-\exp(\rho_2) x_i)] - \sum_{i=1}^{n} \ln[1 - (1 - \exp(\rho_3))(1 - (1 - e^{-\exp(\rho_2) x_i})^{\exp(\rho_1)})] \right\},
$$

and

$$
\pi(\rho_2 \mid \rho_1, \rho_3, x) \propto \exp \left\{ n\rho_1 + n.\ln[\exp(\rho_3) - 1] - n.\ln\rho_3 - (\exp(\rho_1) - 1) \sum_{i=1}^{n} \ln[1 - \exp(-\exp(\rho_2) x_i)] - \sum_{i=1}^{n} \ln[1 - (1 - \exp(\rho_3))(1 - (1 - e^{-\exp(\rho_2) x_i})^{\exp(\rho_1)})] \right\},
$$

and

$$
\pi(\rho_3 \mid \rho_1, \rho_2, x) \propto \exp \left\{ n\rho_1 + n\rho_2 - \sum_{i=1}^{n} \exp(\rho_2) \cdot x_i - (\exp(\rho_1) - 1) \sum_{i=1}^{n} \ln[1 - \exp(-\exp(\rho_2) x_i)] - \sum_{i=1}^{n} \ln[1 - (1 - \exp(\rho_3))(1 - (1 - e^{-\exp(\rho_2) x_i})^{\exp(\rho_1)})] \right\},
$$

Using the WinBUGS software, posterior summaries of interest can be obtained such as the mean; the standard deviation; the credible intervals and others.

A Numerical Example

Consider the data set mentioned in subsection 3.1 We consider the PAL generalized exponential distribution with density (5) under the reparametrization $\rho_1 = \log(\alpha), \rho_2 = \log(\lambda)$, and $\rho_3 = \log(p)$. We assume approximate non-informative prior uniform U(0,1), U(0,0.01) and U(0,0.7) distributions for $\rho_1$, $\rho_2$ and $\rho_3$ respectively.

A set of 10000 Gibbs samples was generated after a “burn-in-sample” of size 1000 to eliminate the initial values considered for the Gibbs sampling algorithm. All the calculations are performed using the WinBUGS software. “One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error (MC error) for each parameter. This is an estimate of the difference between the mean of sampled values and the true posterior mean. The simulation should be run until the MC error for each parameter of interest is less than about
5% of the sample standard deviation”. One can note in our example that MC error less than 5% of the sample standard deviation.

Once the convergence achieved, one need to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference.

The following table lists the posterior descriptive summaries of interest for the PAL generalized exponential model.

Table 2: Summary results for the posterior parameters in the case of the PAL generalized exponential model based on 100 breaking stress data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Deviation</th>
<th>MC error</th>
<th>95% Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>2.713</td>
<td>0.005198</td>
<td>8.43E-5</td>
<td>(2.699, 2.718)</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>1.002</td>
<td>0.001933</td>
<td>3.08E-5</td>
<td>(1.000, 1.007)</td>
</tr>
<tr>
<td>( p )</td>
<td>2.009</td>
<td>0.00441</td>
<td>9.18E-5</td>
<td>(1.998, 2.014)</td>
</tr>
</tbody>
</table>

From Tables 1-2, we conclude that the usual maximum likelihood inference using classical asymptotic results could lead to larger confidence intervals compared to the credible intervals come from the posterior summaries.

References


**Received: November 12, 2016; Published: January 9, 2017**