Research Paper

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

A Statistical Study with Change Point Model in Relation with Water Quality Analysis

STATISTICS

KEYWORDS : Water Quality Analysis, Bayes Estimates, Change Point, Loss Functions, Mixture of Exponential and Degenerate distribution, Reliability Estimation.

Maitreya N. Acharya

Department of Statistics, Maharaja Krishnakumarsinhji Bhavnagar University, Bhavnagar, Gujarat, India.

ABSTRACT

Here, we propose a model with a change point, where the error distribution is suppose to be changing exponentially. Since, the errors are assumed to be distributed as exponential; we term it as a model with "exponential errors". This model can be used as a model for water quality analysis. Later, an independent sequence X1, X2, . . ., Xm, Xm+1, . . . , Xn was observed from the Non-Standard mixture of Exponential and Degenerate Distribution, with reliability R1τ at time t, with proportion p1 and θ1 but later it was found that there was a change in the system at some point of time 'm' and this is reflected in the sequence after Xm by change in reliability R2τ at time t, with proportion p2 and θ2. This distribution occurs in many practical situations. For instance; we consider the statistical analysis of several closely related models arising in water quality analysis, where this model is significantly applicable. Apart from mixture distributions, the phenomenon of change point is also observed in several situations in life testing and reliability estimation problems. The estimators of m, R1 (t) and R2 (t) are derived under Linex Loss Functions & General Entropy Loss Functions. The effects of prior consideration on Bayes Estimates of the change point are also studied.

1. INTRODUCTION:

As far as applications are concerned, we frequently study the time series data. The time series data is to be studied for a variety of different reasons. The time series data have characteristics which are not compatible with the usual assumption of linearity or / and Gaussian errors. For an instance, the exponential model studied by Bell and Smith (1986) which of autoregressive nature is very significant in the analysis of such time series data. Such statistical analysis of several closely related models are useful in study of water quality analysis. Bell and Smith (1986) proposed a model concerned with the above mentioned autoregressive scheme where all the observations were independently and identically distributed and non-negative in nature. The estimation and testing problem was considered by them for Gaussian, Uniform and Exponential models. Due to the additive nature of filtering process, non normality may not be of importance as far as large series is concerned. However, for small series, the effects may be important. Hence, models other than Gaussian, in particular, the exponential, are studied. The Exponential distribution plays an important role in the field of life testing and reliability estimation. A number of life test data have been analyzed (see Davis, 1952) and it was observed that in most cases the exponential distribution provides a good fit.

Let us consider an exponential distribution with probability density function as,

$$f(x|\theta) = \theta^{-1} \exp[-x/\theta], x \geq 0, \theta > 0.$$

The scale parameter θ is the mean residual lifetime and is also interpreted as the average excess life of the items. The model is specified to represent the distribution of lifetimes and statistical inferences are made on the basis of this model. In many real life problems, theoretical or empirical distributions suggest the model with occasionally changing one or more of its parameters.

Let $X_1, X_2,\ldots,X_n$ ($n \geq 3$) be a sequence of observed lifetimes. Let first 'm' observations $X_1, X_2,\ldots,X_m$ be taken from mixture of Exponential and Degenerate Distribution with probability density function as under:

$$f(x;\theta_1,p_1)=(1-p_1)^{I(x_i)}\left[\frac{p_1}{\theta_1}\exp(-\frac{x_i}{\theta_1})\right]^{\left[1-I(x_i)\right]}$$

where we have $\theta_1>0, x_i \geq 0, o< p_1 \leq 1, i=1,2,\ldots,m$

and

$$I(x_i)=\begin{cases}1 & ;x_i<0\\ 0 & ;x_i\geq 0\end{cases}$$

with reliability,

$$R_1(t)=\begin{cases}p_1 & ;t<0\\ p_1\exp[-t/\theta_1] & ;\theta_1>0, t\geq 0\end{cases} \qquad (1)$$

and later (n-m) observations $X_{m+1},X_{m+2},\ldots,X_n$ be taken from the mixture of Exponential and Degenerate distribution with probability density function as

$$f(x;\theta_2,p_2)=(1-p_2)^{I(x_i)}\left[\frac{p_2}{\theta_2}\exp(-\frac{x_i}{\theta_2})\right]^{\left[1-I(x_i)\right]}$$

where we have $\theta_2>0, x_i \geq 0, o< p_2 \leq 1, i=m+1,m+2,\ldots,n,$

with reliability

$$R_2(t)=\begin{cases}p_2 & ;t<0\\ p_2\exp[-t/\theta_2] & ;\theta_2>0, t\geq 0, \theta_1\neq\theta_2\end{cases} \qquad (2)$$

where 'm' is the change point.

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

$$L(\theta_1,\theta_2,m|\underline{x}) = (1-p_1)^{d_m} p_1^{m-d_m} \exp\left[-s_m/\theta_1\right]\theta_1^{-(m-d_m)}$$

$$(1-p_2)^{d_n-d_m} p_2^{(n-m)-d_n+d_m} \exp\left[-\{s_n-s_m\}/\theta_2\right]\theta_2^{-(n-m-d_n+d_m)}$$

(2.5)

where,

$$\sum_{i=1}^{n} I(x_i) = d_n \quad ; \quad \sum_{i=1}^{m} I(x_i) = d_m \quad ; \quad \sum_{i=m+1}^{n} I(x_i) = d_n - d_m$$

$$\sum_{i=1}^{n} [x_i(1-I(x_i))] = s_n \; ; \; \sum_{i=1}^{m} [x_i(1-I(x_i))] = s_m \; ; \; \sum_{i=m+1}^{n} [x_i(1-I(x_i))] = s_n - s_m .$$

3. POSTERIOR DENSITIES USING NON-INFORMATIVE PRIOR:

Sometimes no prior information or technical knowledge about the parameters are available.

In such cases, we take non-informative priors.

Let us consider such non-informative prior on θ_1,θ_2 as,

$$g(\theta_1,\theta_2) \propto 1/(\theta_1\theta_2).$$

Then the joint prior distribution of θ_1,θ_2 and m is,

$$g(\theta_1,\theta_2,m) \propto 1/(\theta_1\theta_2(n-1)). \qquad (3)$$

Using the likelihood function with the prior density, we get the joint posterior density as

under:

$$g_2[\theta_1,\theta_2,m|\underline{x}] = \frac{L(\theta_1,\theta_2,m|\underline{x})g(\theta_1,\theta_2,m)}{\sum_{m=1}^{n-1}\int_0^\infty\int_0^\infty L(\theta_1,\theta_2,m|\underline{x})g(\theta_1,\theta_2,m)d\theta_1 d\theta_2}$$

Now the marginal density of $\underline{x}$, $h_2(\underline{x})$ is simplified as,

$$h_2(\underline{x}) = \sum_{m=1}^{n-1}\int_0^\infty\int_0^\infty L(\theta_1,\theta_2,m|\underline{x})g(\theta_1,\theta_2,m)d\theta_1 d\theta_2$$

$$= (n-1)^{-1}\sum_{m=1}^{n-1} p_1^{m-d_m}(1-p_1)^{d_m}\, p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}$$

$$\left\{\int_0^\infty \theta_1^{-(m-d_m)}\exp[-s_m/\theta_1]d\theta_1 \int_0^\infty \theta_2^{-(n-m-d_n+d_m)}\exp[-\{s_n - s_m(n-m-d_n+d_m)\}/\theta_2]d\theta_2\right\}$$

$$= (n-1)^{-1}\sum_{m=1}^{n-1} p_1^{m-d_m}(1-p_1)^{d_m}\, p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}\,\overline{\lfloor}(m-d_m)\,\overline{\lfloor}(n-m-d_n+d_m+a_2)A_2^{-(m-d_m)}B_2^{-(n-m-d_n+d_m)}$$

$$= (n-1)^{-1}\sum_{m=1}^{n-1}(k_4(m)J_2(m)). \qquad (4)$$

where,

$$A_2 = A_2(m|\underline{x}) = s_m,$$

$$B_2 = B_2(m|\underline{x}) = s_n - s_m, \qquad (5)$$

$$k_4(m) = \overline{\lfloor}(m-d_m)\,\overline{\lfloor}(n-m-d_n+d_m)\; p_1^{m-d_m}(1-p_1)^{d_m}\, p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}$$

$$J_2(m) = A_2^{-(m-d_m)}B_2^{-(n-m-d_n+d_m)}$$

Hence with the use of these, we get the joint posterior density,

$$g_2[\theta_1,\theta_2,m|\underline{x}]$$

$$= (n-1)^{-1}\, p_1^{m-d_m}(1-p_1)^{d_m}\, p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}\,\theta_1^{-(m-d_m)-1}\exp[-A_2/\theta_1]$$

$$\theta_2^{-(n-m-d_n+d_m)-1}\exp[-B_2/\theta_2]\,[h_2(\underline{x})]^{-1}\ . \qquad (6)$$

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

Marginal posterior distribution of the change point m, say, $g_2(m \mid \underline{x})$, is given by

$$g_2(m \mid \underline{x}) = k_4(m) \cdot J_2(m) / \sum_{m=1}^{n-1} \left(k_4(m) \cdot J_2(m)\right) \qquad (7)$$

where, $k_4(m)$ and $J_2(m)$ have same meanings as above.

Here, A_2, B_2 have same meanings as in the above equation and $h_2(\underline{x})$ also has the same meaning as mentioned in the above equation.

The marginal posterior distribution on θ_1, say $g_2\left[\theta_1 \mid \underline{x}\right]$ is,

$$g_2\left[\theta_1 \mid \underline{x}\right] = \frac{(n-1)^{-1}}{h_2(\underline{x})} p_1^{m-d_m} (1-p_1)^{d_m} p_2^{n-m-d_n+d_m} (1-p_2)^{d_n-d_m} \theta_1^{-(m-d_m)-1} \exp\left[-A_2/\theta_1\right]$$

$$\int_0^\infty \theta_2^{-(n-m-d_n+d_m)} \exp\left[-B_2/\theta_2\right] d\theta_2$$

$$= \frac{(n-1)^{-1}}{h_2(\underline{x})} p_1^{m-d_m} (1-p_1)^{d_m} p_2^{n-m-d_n+d_m} (1-p_2)^{d_n-d_m} \Gamma\left(n-m-d_n+d_m\right) \theta_1^{-(m-d_m)-1} \exp\left[-A_2/\theta_1\right] \frac{1}{B_2^{\,n-m-d_n+d_m}} \qquad (8)$$

$$g_2\left[\theta_2 \mid \underline{x}\right] = \frac{(n-1)^{-1}}{h_2(\underline{x})} p_1^{m-d_m} (1-p_1)^{d_m} p_2^{n-m-d_n+d_m} (1-p_2)^{d_n-d_m} \theta_2^{-\left(n-m-d_n+d_m\right)-1} \exp\left[-B_2/\theta_2\right]$$

$$\int_0^\infty \theta_1^{-(m-d_m)} \exp\left[-A_2/\theta_1\right] d\theta_1$$

$$= \frac{(n-1)^{-1}}{h_2(\underline{x})} p_1^{m-d_m} (1-p_1)^{d_m} p_2^{n-m-d_n+d_m} (1-p_2)^{d_n-d_m} \theta_2^{-\left(n-m-d_n+d_m\right)-1} \exp\left[-B_2/\theta_2\right] \Gamma\left(m-d_m\right) \frac{1}{A_2^{\,m-d_m}}$$

(9)

Making the change of variables and taking i=1, 2, the marginal posterior densities on the reliability functions $R_1(t_0)$ and $R_2(t_0)$ will be as under:

$$g_2\left[r_1(t_0)\middle|\underline{x}\right]=\frac{(n-1)^{-1}p_1^{-1}}{h_2(\underline{x})}\sum_{m=1}^{n-1}k_5(m)\left[\ln(p_1/r_1(t_0))\right]^{m-d_m-1}$$

$$\frac{\left[r_1(t_0)/p_1\right]^{(A_2/t_0)-1}}{(t_0)^{m-d_m}B_2^{n-m-d_n+d_m}}\qquad(\text{here } 0<R_1(t_0)<p_1) \tag{10}$$

and

$$g_2\left[r_2(t_0)\middle|\underline{x}\right]=\frac{(n-1)^{-1}p_2^{-1}}{h_2(\underline{x})}\sum_{m=1}^{n-1}k_6(m)\left[\ln(p_2/r_2(t_0))\right]^{n-m-d_n+d_m-1}$$

$$\frac{\left[r_2(t_0)/p_2\right]^{(B_2/t_0)-1}}{(t_0)^{n-m-d_n+d_m}A_2^{m-d_m}}\quad(\text{here } 0<R_2(t_0)<p_2) \tag{11}$$

where,

$$k_5(m)=p_1^{m-d_m}(1-p_1)^{d_m}\,p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}\,\overline{\ }\,(n-m-d_n+d_m),$$

$$k_6(m)=p_1^{m-d_m}(1-p_1)^{d_m}\,p_2^{n-m-d_n+d_m}(1-p_2)^{d_n-d_m}\,\overline{\ }\,(m-d_m), \tag{12}$$

where, $E_2=\min\{c_2, t_0\}$.

4. BAYES ESTIMATES UNDER SYMMETRIC LOSS FUNCTIONS:

The Bayes estimate of a generic parameter(or function thereof) α based on a Squared Error Loss (SEL) function $L_1(\alpha,d)=(\alpha-d)^2$,where, d is decision rule to estimate α, is the posterior mean. As a consequence, the Square Error Loss Function is related to an integer parameter,

$$L_1'(m,v)\propto(m-v)^2,\qquad m,v=0,1,2...$$

Hence, the Bayesian estimate of an integer-valued parameter under the SEL function $L_1'(m,v)$ is no longer the posterior mean and can be obtained by numerically minimizing the corresponding

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

posterior loss. Generally, such a Bayesian estimate is equal to the nearest integer value to the posterior mean. So, we tell the nearest value to the posterior mean as Bayes Estimate.

The Bayes estimate of unknown change point 'm', using non-informative priors, discussed earlier under SEL is given by,

$$m^{*} = \sum_{m=1}^{n-1} m \cdot g_1(m|\underline{x})$$

$$m^{*} = \sum_{m=1}^{n-1} \left(m \cdot k_1(m) J_1(m) / \sum_{m=1}^{n-1} k_1(m) J_1(m) \right) \qquad (13)$$

Let $E_i\{f(\alpha)\}$ denote the expectation of f (α) with respect to the posterior density $g_i(\alpha|\underline{x})$ where i=1, 2. Then Bayes estimates of $R_1(t_0)$ and $R_2(t_0)$ under SEL, are posterior means $R_1^{*}(t_0)$ and $R_2^{*}(t_0)$ given by,

$$R_1^{*}(t_0) = E_1[R_1(t_0)]$$

$$= \int_0^{p_1} r_1(t_0)\ g_1[r_1(t_0)|\underline{x}]\ dr_1(t_0).$$

Now,

$$\int_0^{p_1} r_1(t_0)\ g_1[r_1(t_0)|\underline{x}]\ dr_1(t_0)$$

$$= \frac{k_0}{h_1(\underline{x})} \sum_{m=1}^{n-1} k_2(m) \frac{1}{(t_0)^{m-d_m+a_1} B_1^{n-m-d_n+d_m+a_2}}$$

$$\left\{ \int_0^{p_1} \left[r_1(t_0)/p_1 \right] (A_1/t_0)\ \left[\ln(p_1/r_1(t_0)) \right]^{m-d_m+a_1-1} dr_1(t_0) \right\}$$

$$= \frac{k_0\ p_1}{h_1(\underline{x})} \sum_{m=1}^{n-1} k_2(m) \frac{1}{(A_1+t_0)^{m-d_m+a_1} B_1^{n-m-d_n+d_m+a_2}}$$

Hence,

$$R_1^*(t_0)$$

$$= \left\{ \frac{k_0 p_1}{h_1(\underline{x})} \sum_{m=1}^{n-1} k_1(m) \frac{1}{(A_1 + t_0)^{m - d_m + a_1} \; B_1^{n-m-d_n + d_m + a_2}} \right\}, \tag{14}$$

Similarly,

$$R_2^*(t_0) = E_1[R_2(t_0)]$$

$$= \int_0^{p_2} r_2(t_0) \;\; g_1[r_2(t_0)|\underline{x}] \;\; dr_2(t_0) .$$

$$= \left\{ \frac{k_0 \;\; p_2}{h_1(\underline{x})} \sum_{m=1}^{n-1} k_1(m) \frac{1}{A_1^{m - d_m + a_1} (B_1 + t_0)^{n-m-d_n + d_m + a_2}} \right\} \tag{15}$$

where, $\Pr\{R_i(t_0) = 1|\underline{x}\}$, i=1,2 are given above and $k_1(m)$ and $k_2(m)$ respectively are same as mentioned earlier.

Other Bayes estimators of α based on the loss functions

$$L_2(\alpha, d) = |\alpha - d|$$

$$L_3(\alpha, d) = \begin{cases} 0, & if \; |\alpha - d| < \varepsilon, \; \varepsilon > 0 \\ 1, & otherwise \end{cases}$$

is the posterior median and posterior mode respectively.

Volume : 5 | Issue : 5 | May 2016 • ISSN No 2277 - 8179 | IF : 3.508 | IC Value : 69.48

5. NUMERICAL STUDY (when p_1 and p_2 are known):

We have generated 20 random observations from the proposed change point model explained earlier, the first ten observation from mixture of exponential and degenerate distribution with $\theta_1 = 0.7$, $R_{1\tau}$=0.78, p_1=0.8 at t=0.01 and next ten from mixture of exponential and degenerate distribution with $\theta_2 = 0.5$, $R_{2\tau}$=0.58, p_2=0.6 at t=0.01. As explained above, θ_1 and θ_2 themselves were random observations with means $\mu_1 = 0.7$, $\mu_2 = 0.5$ and standard deviations σ_1=1 and σ_2=2 respectively, resulting in a_1=2.49, b_1=1.04, a_2=2.06, b_2=0.53. These observations are given below in Table 1.

Table 1

Generated observations from Mixture of Exponential and Degenerate Distribution.

I	1	2	3	4	5	6	7	8	9	10
X_i	0.229	1.426	0	1.103	0.199	0.226	0.948	0.482	0	0.283
I	11	12	13	14	15	16	17	18	19	20
X_i	0	1.018	0.187	0.788	0	0.161	0.677	0	0.331	0

Table-2

Bayes estimates using symmetric loss function

Prior Density	Shape parameter		Bayes estimates of Change point		Bayes estimates of Reliability under Squared Error Loss	
	q_1	q_2	m_L^*	m_{GE}^*	$R_{1E}^*(t)$	$R_{2E}^*(t)$
Non-Informative	-1.0 -2.0	-1.0 -2.0	11.0 12.0	12.0 13.0	0.82 0.84	0.83 0.85

6. CONCLUSION:

Our numerical study shows that for $q_1 = q_3 = -1, -2$, Bayes estimates are quite large than actual value, ie; m=10. It can be seen from Table 2 that if we take the value of shape parameters of loss function negative, underestimation can be solved.

Moreover, Table 2 shows that, for small values of $|q|, q_3$ related to the Squared Error Loss Function, the values of Bayes Estimate under a loss is near by posterior mean. Table 2 also shows that, for $q_2 = 1.5, 1.2$, Bayes Estimates are less than actual value ie; m=10.

REFERENCES

1. Aitchison, J. (1955). On the distribution of positive random variable having a discrete probability mass at the origin. J. Amer.Stat Assn. Vol. 50,901-908
2. Calabria, R. and Pulcini, G. (1994c). "Bayes credibility intervals for the left-truncated exponential distribution", Micro electron Reliability, 34, 1897-1907.
3. Calabria, R. and Pulcini, G. (1996). "Point estimation under asymmetric loss functions for left-truncated exponential samples", Communication in Statistics (Theory and Methods), 25(3), 585-600.
4. Davis, D. J. (1952). "The analysis of some failure data", J. Am. Statist. Assoc., 47, 113-160.
5. Ebrahimi N. and Ghosh S. K. (2001). Bayesian and frequentist methods in change-point problems. Handbook of statistic, Vol.20, Advance in Reliability, 777-787, (Eds. N. Balakrishna and C. R. Rao).
6. Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change point in a sequence of binomial random variables. Biometrika, 57, 477-488.
7. Jani, P. N. and Pandya, M. (1999). Bayes estimation of shift point in left Truncated Exponential Sequence Communications in Statistics (Theory and Methods), 28(11), 2623-2639.
8. Kander, Z. and Zacks, S. (1966). Test procedure for possible changes in parameters of statistical distribution occurring at unknown time points. Annals Math. Statist., 37, 1196-1210.
9. Muralidharan, K. (1999). Tests for the mixing proportion in the mixture of a Degenerate and Exponential distribution. J. Indian Stat. Assn., Vol. 37, Issue 2.
10. Pandya, M. and Jani, P.N. (2006). Bayesian Estimation of Change Point in Inverse Weibull Sequence. Communication in statistics- Theory and Methods, 35 (12), 2223-2237.
11. Pandya, M. and Jadav, P. (2007). Bayesian Estimation of Continuous Change Point in Inverse Weibull distribution. Int. J. Agricultural Stat. Sci., 3 (2), 589-595.
12. Pandya M. and Krishnam Bhatt (2010) Application of AR (1) with change point on total population of India from official statistics, Census of India.
13. Smith, A. F. M. (1975). A Bayesian approach to inference about a change point in a sequence of random variables. Biometrika, 62, 407-416.
14. Vannman. K. (1991). Comparing samples from nonstandard mixture of distributions with applications to quality comparison of wood. Research report 1991:2 submitted to Division of Quality Technology, Lulea University, Lulea, Sweden.
15. Vannman, K. (1995). On the distribution of the estimated mean from the nonstandard Mixtures of distribution. Comm. Statist. – Theory and Methods, 24(6), 1569-1584.
16. Zacks, S. (1983). Survey of Classical and Bayesian approaches to the change point problem: fixed sample and sequential procedures for testing and estimation. Recent advances in statistics. Herman Chernoff Best Shrift, Academic Press New-York, 1983, 245-269.