Science World Journal Vol 13(No 3) 2018
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University.

Full Length Research Article

THE USE OF GOMBAY METHOD IN MONITORING OBSERVATIONS WITH WEIBULL DISTRIBUTION

Samson Henry Dogo

Department of Mathematical Sciences, Kaduna State University, Kaduna State – Nigeria.

*Author Email Address: samson.dogo@kasu.edu.ng

ABSTRACT
This paper presents a sequential test for monitoring observations with a Weibull Distribution based on Gombay method. The Gombay method is a sequential method for parametric models in the presence of a nuisance parameter. The scale parameter of the Weibull is considered as the parameter of interest λ while the shape parameter as the nuisance parameter. Simulations shows that the Type I error rate decrease with increase in sample size K and shape Parameter $\propto$. The Type I errors achieved in all combination of the variables considered is below the nominal level of 0.05, which makes it conservative.

Keywords: Weibull distribution, Sequential analysis and Gombay

1. INTRODUCTION

Sequential analysis was initially introduced during World War II in response to the overwhelming demands for methods of testing efficiency of aircraft gunnery (Lai, 2001). In this method of analysis the sample size is not fixed in advance, instead data are evaluated as more observations are collected, and further sampling is stopped in accordance with a predefined stopping rule as soon as significance results is observed. Sequential methods are increasingly becoming popular statistical methods used for monitoring mechanical components, quality assurance, and quality of medical services such as the effectiveness of clinical procedure, disease epistemology and patients recovery in hospital. For example in controlled clinical trials, sequential methods are used for gauging sufficiency in treatment effect (Lau et al, 1992; Pogue and Yusuf, 1997; Wetterslev et al, 2008) and serve as an appropriate statistical tool for identifying when evidence becomes significantly harmful.

The Page (1954) cumulative sum scheme (CUSUM) is the most widely used sequential methods for quality control and monitoring of production process in industries. This is due to its ability to detect small changes and incorporate all information in the sequence of values in the sampled data (Montgomery, 2000). However, the design of CUSUM scheme requires that the sequence of observations be independent and identically distributed, which may not be true in most practical problems as in meta-analysis (Dogo et al, 2015). Another problem is that in most sequential methods for Weibull the shape parameter is usually assumed or held fixed for every sequence of observations (see example in Steiner et al., 2000), which is not true in practice. However, the extension of the CUSUM scheme developed by Gombay (2003), and Gombay and Serbian (2005) has the CUSUM properties and useful in monitoring parametric models in the presence of a nuisance parameter- a parameter that is not of immediate interest but most be accounted for in the course of analysis. Henceforth, this method shall be referred to as Gombay method throughout the paper. In this method there is no need to assumed or held the shape parameter fixed, rather it is considered as a nuisance parameter.

The objective of this paper is to define a Gombay test for monitoring sequential observations with Weibull distribution. The Weibull distribution has many applications that include survival analysis, weather forecasting and in hydrology it is used to study extreme events such as modelling the distribution of annual rainfall. The rest of the paper is organised as follows. In Section 2, we present the general Gombay method. Section 3 stipulates the Gombay method for monitoring observation with Weibull distribution. Section 4 gives a simulation study on Gombay method for observations with Weibull distribution. Section 5 is the summary and conclusion.

2. GOMBAY METHOD

Before presenting the Gombay method we briefly introduce the score test which is closely related to the Gombay method.

2.1. Score Test

The score test introduced by Rao (1948) is a fixed sample size test of a null hypothesis that a parameter of interest takes a particular value. Let X be an independent random variable with density $f(X;\ \omega)$, where ω is a parameter of interest. The score statistic for testing a null hypothesis, $H_0\colon \omega = \omega_0$ is defined (Rao, 1948) by

$$S(\omega) = \frac{(\mu(\omega))^2}{I(\mu)}, \qquad (1)$$

Where $\mu(\omega) = \frac{\partial}{\partial\omega} log f(X,\omega)$ is known as the efficient score vector and $I(\omega) = var\big(\mu(\omega)\big) = E_X[(\mu(\omega))^2] = E_X\left[\left(\frac{\partial}{\partial\omega} log f(X,\omega)\right)^2\right]$ is the Fisher information and the derivatives are taken at ω_0. Under the null hypothesis the statistic $S(\omega)$ is χ^2 distributed with 1 degrees of freedom (Rao, 1948).

In most statistical problems, ω is rarely only a parameter of interest. Let $\omega = (\theta,\eta)$, such that the observed variable $X \sim f(.\,;\theta,\eta)$, θ is a vector of real parameters of interest and η is the vector of nuisance parameters. Since the interest is in inference about θ, it is important to find a way to deal with the nuisance parameter. One way to do this is to eliminate the nuisance parameter by conditioning the score statistic, (see (Lindsey, 1983; Basu, 1977)). A suitable statistic is chosen, say $g(x\epsilon X)\colon (x\epsilon X,\omega) \to (y\epsilon Y)$ such that the conditional distribution of μ^c depends on ω only through θ. The conditional score vector may be defined (Lindsey, 1983) by

$$g(x\epsilon X) = \mu(\omega) - E[\mu/T], \qquad (2)$$

where T is a sufficient statistic whose sampling distribution depends on θ only. If θ is real-valued, the information corresponding to $g(x\epsilon X)$ is obtained from a Fisher information matrix for the parameter $(\theta,)$ given (Gombay and Serbian, 2005;

Lindsey, 1983) by

$$I = \begin{pmatrix} I_{\theta\theta} & I_{\theta\eta} \\ I_{\eta\theta} & I_{\eta\eta} \end{pmatrix}, \quad (3)$$

And the marginal information about θ, also known as the effective information is given by $I(\theta) = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}$, see (Bera and Bilias, 2001; Gombay and Serbian , 2005)

2.2. Sequential Hypotheses and Gombay Test Statistic

While the score test is a fixed sample test of a null hypothesis that a parameter of interest takes a particular value, the Gombay test is a sequential change detection test with the test statistic defined by the maximum of a sequence of score tests, $S_j = S: S_{(j)} = c(X_1, X_2, \dots, X_j)$ calculated from the sequence of observed data, $G_k = max\{S_1, S_2, \dots, S_j\}$. Below is the description of Gombay method introduced as test I in Gombay and Serbian (2005). Consider a sequence of independent random variables, $X_1, X_2, \dots . \sim f(., \theta_i, \eta_i)$, where f is a probability density function, θ is a vector of parameter of interest and η is a nuisance parameter. Consider a test for the composite hypotheses

$$H_0: \theta_i = \theta_0, \ \eta_i = \eta; \ i = 1, 2, \dots.$$

Against and alternative

$$H_1: \begin{cases} \theta_i = \theta_0, \ \eta_i = \eta; \ i = 1, 2, \dots, r \\ \theta_i = \theta_0 + \Delta\theta, \ \eta_i = \eta; \ i \geq r + 1, \end{cases}$$

where $r \geq 1$ is an unknown time of change, $\Delta\theta$ is a shift in the value of the parameter of interest from θ_0 and η an unknown nuisance parameter. The null value of the vector of parameter of interest θ_0 can take any value from R^d.

Let $\omega = (\theta, \eta)$, the log-likelihood function at $k - th$ interim analysis is $l(\omega) = \sum_{i=1}^{k} lnf(X_i, \omega)$. And the efficient score vector for θ and η is defined by

$$V_k(\theta_0, \eta) = \frac{\partial l(\omega)}{\partial \omega} = \sum_{i=1}^{k} \frac{\partial}{\partial \omega} lnf_{\theta_0\eta}(X_i). \quad (4)$$

In order to define a test statistic for the hypotheses about θ, a Fisher information matrix I for k observations is partition as in equation (3), where $I_{\theta\theta} = -E\left(\frac{\partial^2}{\partial\theta^2} l(\theta, \eta)\right)$, $I_{\eta\eta} = -E\left(\frac{\partial^2}{\partial\eta^2} l(\theta, \eta)\right)$ and $I_{\theta\eta} = I_{\eta\theta} = -E\left(\frac{\partial^2}{\partial\theta\partial\eta} l(\theta, \eta)\right)$.

Replacing the nuisance parameter η with its maximum likelihood estimate $\hat{\eta}_k$, obtained from the solution of

$$\sum_{i=1}^{k} \frac{\partial}{\partial \eta} lnf_{\theta_0\eta}(X_i) = 0, \quad (5)$$

the conditional efficient score vector V_k is given by

$$V_k(\theta_0, \hat{\eta}_k) = \sum_{i=1}^{k} \frac{\partial}{\partial \eta} lnf_{\theta_0\eta}(X_i). \quad (6)$$

This vector is sometimes termed effective score vector and its variance $\Gamma_k = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}$ is called effective information (Bera and Billias, 2001). For independent random variables the variance increases linearly with numbers of observations, i.e. $\Gamma_k(\theta_0, \eta) = k\Gamma_1(\theta_0, \eta)$.Under some standard regularity conditions guaranteeing the existence and consistence of a sequence of maximum likelihood estimates given by Serfling (1980) and Lehmann (2001), Gombay and Serbian (2005) showed that under H_0, as $k \to \infty$, the effective score vector can be written as

$$V_k(\theta_0, \hat{\eta}_k) = \sum_{i=1}^{k} \frac{\partial}{\partial \eta} lnf_{\theta_0\eta}(X_i)$$

$$= \sum_{i=1}^{k} \left\{\frac{\partial}{\partial \theta} lnf_{\theta_0\eta}\right\}$$

$$= \sum_{i=1}^{k} \left\{\frac{\partial}{\partial \eta} lnf_{\theta_0\eta} I_{\theta\theta}(\theta_0, \eta) - I_{\theta\eta}(\theta_0, \eta) I_{\eta\eta}^{-1}(\theta_0, \eta) I_{\eta\theta}(\theta_0, \eta)\right\} + O(loglogk)$$

$$= \sum_{i=1}^{k} Z_i + O(loglogk),$$

where Z_i are independent identically distributed random variables with expected value $E[Z_i] = 0$ and covariance $cov(Z_i) = k^{-1}\Gamma_k(\theta_0, \eta)$, for $\Gamma_k = I_{\theta\theta} - I_{\theta\eta} I_{\eta\eta}^{-1} I_{\eta\theta}$. It follows that

$$T_k = \sqrt{k}\Gamma_k(\theta_0, \eta)^{-1/2} \sum_{i=1}^{k} lnf_{\theta_0\eta}(X_i) \quad (7)$$

is asymptotically $(k \to \infty)$ the sum of independent and identical random variables with mean 0 and variance 1, and thus a sequence of statistics $\{T_k\}$ can be approximated by the standard Weiner process. In order to use the statistic T_k, for testing, the covariance $\Gamma_k(\theta_0, \eta)$ is replaced with its estimate $\Gamma_k(\theta_0, \hat{\eta}_k)$. Gombay (2003) and Gombay and Serbian (2005) introduced a sequential change detection test based on the statistic T_k in (7) as follows. For $k = 1, 2, \dots, K$, where K is a truncation point, reject H_0 in favour of a positive change $\Delta\theta > \theta$ at time k if

$$G(K) = \max_{1 < k \leq K} \left\{\frac{1}{\sqrt{K}} T_k\right\} \geq \sqrt{K} C(\alpha), \quad (8)$$

and if no such k, $k \leq K$ exists do not reject H_0. See Gombay (2003) and Gombay and Serbian (2005) for asymptotic critical values $C(\alpha)$, and detailed derivation and discussion of the method.

3. Gombay Test for Observations with Weibull Distribution

To define the Gombay test for the Weibull, it suffices to determine the efficient score vector, elements of Fisher information matrix as defined in Section 2.2. Suppose $x_t, for \ t = 1, 2, \dots.,$ are sequence of observations that follows the Weibull, then the probability density function of the distribution of the observation can be defined by

$$f(x_t) = \frac{\alpha}{\lambda}\left(\frac{x_t}{\lambda}\right)^{\alpha-1} exp\left\{-\left(\frac{x_t}{\lambda}\right)^{\alpha}\right\}; \ x \geq 0, \alpha, \lambda > 0 \quad (9)$$

where x_t is the observed data at time t, α is the shape parameter and λ is the scale parameter that indicates the spread of the distribution of the observed data. In sequential methods for Weibull, the scale parameter is the parameter of interest since it depends on the covariate, and therefore we choose λ as the parameter of interest and α is the nuisance parameter. Let $\lambda = \lambda_0$ be the null value of the parameter of interest. As more observations are taken, the goal is to determine when the value of the scale parameter changes significantly from the null value λ_0, if at all, and stop further sampling.

The log-likelihood function of the Weibull distribution is given by

$$L(\alpha, \lambda) = nln\alpha - n\alpha ln\lambda - \frac{1}{\lambda^2} \sum_{i=1}^{n} x^{\alpha} + (\alpha - 1) \sum_{i=1}^{n} lnx. \quad (10)$$

To obtain the maximum likelihood estimates of the Weibull parameters, differentiating equation (10) with respect to α and λ and equating to zero, the following equations were obtained.

$$\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} - nln\lambda - \frac{\sum_{i=1}^{n} x_i^{\alpha} lnx_i - ln\lambda \sum_{i=1}^{n} x_i^{\alpha}}{\lambda^{\alpha}} + \sum_{i=1}^{n} lnx_i = 0 \quad (11)$$

$$\frac{\partial L}{\partial \lambda} = -\frac{n}{\lambda}\alpha + \frac{\alpha}{\lambda^{\alpha-1}} \sum_{i=1}^{n} x_i^{\alpha} = 0 \quad (12)$$

Solving for λ in equation (12) and substituting the result in (11) the results are given by

$$\lambda = \left[\frac{1}{n} \sum_{i=1}^{n} x_i^{\alpha}\right]^{1/\alpha} \quad (13)$$

$$1/\alpha - \frac{\sum_{i=1}^{n} x_i^{\alpha} lnx_i}{\sum_{i=1}^{n} x_i^{\alpha}} + 1/n \sum_{i=1}^{n} lnx_i = 0. \quad (14)$$

The maximum likelihood estimate of the parameter α is calculated by solving equation (14) numerically using Newton-Raphson method. The estimate of λ is obtained by using the solution of equation (14) to solve (13).

It follows from equations (6) and (12) that the efficient score vector is defined by

$$V_k(\lambda_0, \hat{\alpha}_k) = -\frac{k}{\lambda_0}\hat{\alpha}_k + \frac{\hat{\alpha}_k}{\lambda_0^{\hat{\alpha}_k - 1}}\sum_{t=1}^{k} x_t^{\hat{\alpha}_k}. \quad (15)$$

To obtain the elements of the Fisher information, we use the formulas presented in Section 2.2, and all terms were approximated by their 2-terms series expansion of the functions before taking the expectations. The results obtain are as follows.

$$I_{\lambda\lambda} = -\hat{\propto}\, k\{\lambda_0^{-2} + (1 - \hat{\propto})\lambda_0^{-\alpha}\}\left\{1 + \hat{\propto}\left(\Gamma\left(1 + {}^{1}/_{\hat{\propto}}\right) - 1\right)\right\} \quad (16)$$

$$I_{\alpha\lambda} = I_{\lambda\alpha} = \frac{k}{\lambda_0} - k\left\{\begin{array}{c} \lambda_0^{1-\hat{\propto}}\left[\lambda_0^{\hat{\propto}}\left(1 + \hat{\propto}\left(\Gamma\left(1 + \frac{1}{\hat{\propto}}\right) - 1\right)\right)\right] \\ + \hat{\propto}\,\lambda_0\left[log\lambda_0 + (\hat{\propto}\, log\lambda_0 + 1)\left(\Gamma\left(1 + \frac{1}{\hat{\propto}}\right) - 1\right)\right] \\ - \hat{\propto}\,\lambda_0 log\lambda_0[1 + \hat{\propto}\,(\Gamma(1 + 1/\hat{\propto}) - 1)] \end{array}\right\} \quad (17)$$

$$I_{\alpha\alpha} = \frac{k}{\hat{\propto}^2} + \frac{k}{\lambda_0}\left\{\begin{array}{c} \lambda_0^{\hat{\propto}} log^2\lambda_0 + \lambda_0^{\hat{\alpha}} log\lambda_0[\hat{\propto}\, log\lambda_0 + 2]\left[\lambda_0\Gamma\left(1 + \frac{1}{\hat{\propto}}\right) - 1\right)\right] \\ + \lambda_0^{2} log^2\lambda_0\left[1 + 2\left(\Gamma\left(1 + \frac{1}{\hat{\propto}}\right) - 1\right)\right] \\ -2\lambda_0^{2} log\lambda_0\left[log\lambda_0 + (1 + 2log\lambda_0)\left(\Gamma\left(1 + \frac{1}{\hat{\propto}}\right) - 1\right)\right] \end{array}\right\} \quad (18)$$

Denote $\xi_k = I_{\lambda\lambda} - I_{\lambda\alpha}I_{\alpha\alpha}^{-1}I_{\alpha\lambda}$ the estimate of Fisher information at the $k - th$ interim analysis, the Gombay test statistic is based on the maximum of the standardized and scaled by $\sqrt{k}$ score statistic (8) given by

$$T_k = \sqrt{k}\,\frac{-\frac{k}{\lambda_0}\hat{\alpha}_k + \frac{\hat{\alpha}_k}{\lambda_0^{\hat{\alpha}_k - 1}}\sum_{t=1}^{k} x_t^{\hat{\alpha}_k}.}{\xi_k}. \quad (19)$$

4. The Simulation Study

The objective of the simulation study presented here is to evaluate the Type I error rate of the Gombay test for observations with Weibull distribution with standard critical values in relation to the sample size K and the shape parameter $\propto$. The data for the simulation were generated from the Weibull distribution, $x_i \sim wei(K, \propto, \lambda_0)$, where λ_0 is the null value of the scale parameter. Critical values were calculated based on 5% significance level and the null value of the parameter of interest, the scale parameter set at $\lambda_0 = 0.5$. Sequential testing starts with a minimum of two observations and stops as soon as a boundary value is reached or after the $K - th$ interim analysis. For each combination of the following variables: $\lambda_0 = 0.5$, $\alpha = (0.6,\ 0.8,\ 1.0)$ and $K = (20,\ 40,\ 60,\ 80)$. A total of 1000 simulations were conducted, the empirical power of the test to reject H_0 was calculated and recorded. The values of the variables used are intentionally chosen to reflect most practical situations.

Table I: Relationship between sample size K, shape parameter $\propto$ and the Type I error rate of Gombay test for observation with Weibull distribution, λ_0

K	20	20	20	40	40	40	60	60	60	80	80	80
$\propto$	0.6	0.8	1.0	0.6	0.8	1.0	0.6	0.8	1.0	0.6	0.8	1.0
Type I error	$7e^{-4}$	$5e^{-4}$	$4e^{-4}$	e^{-3}	$6e^{-4}$	$5e^{-4}$	$4e^{-4}$	e^{-4}	e^{-4}	$3e^{-4}$	$4e^{-4}$	$2e^{-4}$

Achieved Type I error is an important tool for evaluating the quality of a statistical hypothesis test (Dogo et al., 2015). Table I shows the Type I error achieved by the Gombay test for observations with Weibull distribution. Observed that the Type I error decrease as the shape parameter $\propto$ and sample size K increases. The Type I errors achieved in all combinations of the variables considered is below the nominal level of 0.05, which shows that the test is conservative.

5. Summary and Conclusion

The objective of this paper was to define a Gombay test for monitoring observations with Weibull distribution. This has been achieved by setting the scale parameter as the parameter of interest and the shape parameter as the nuisance parameter. Simulations shows that the Type I error achieved is below the nominal level, and are affected by increase in sample size and shape parameter. Therefore, the test is not recommended for use in practice.

The lack of control of the Type I error may be explained by approximation used in taking expectation in the calculations of the elements of Fisher information. Another problem is the use of asymptotic approximation based on Wiener's process to obtain the critical values of the test. However, it will not be self-serving to say that the Gombay method provides a basis for sequential approach to observations with Weibull distribution can based upon. This will be pursue in our future research.

REFERENCES

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of American Statistical Association, 72(358):355-366.*

Bera, A. and Bilias, Y. (2001). Raos Scores, Neymans $C(\propto)$ and Silveys LM tests. An essay on historical developments and some new results. *Journal of Statistical Planning and Inference, 97:9-44.*

Dogo, S. H., Clark, A., and Kulinskaya, E. (2015). A Sequential approach for random-effects meta- analysis. *International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering, 9(1).*

Gombay, E. (2003). Sequential change –point detection and estimation. *Sequential Analysis, 22(3): 203-222.*

Gombay, E. and Serbian, D. (2005). An adaptation of Pages CUSUM test for change detection. *Periodica Mathematica Mathematica Hungarion, 50(1): 135-147.*

Lai, T. L. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, *11(2):303-350*

Lau, J., Antman, E., Jimenez-Silva, J., Kupelnick, B., Mosteller, F., and Chalmers, T. (1992). Cumulative meta- analysis of therapeutic trials for myocardial infarction. *New England Journal of Medicine, 327(4): 248-254.*

Lehmann, E. (2001). Elements of large sample theory.

Lindsey, B. (1983). Efficiency of the conditional score in a mix setting. *The annals of statistics. Pages 486-497.*

Page, E. (1954). Continuous inspection scheme. *Biometrika, 14:100-115.*

Pogue, J. and Yusuf, S. (1997). Cumulative evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Controlled Clinical trials, 18(6): 580-593.*

Rao, C. R., (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *In Mathematical Proceedings of the Cambridge Society, volume 44, pages 50-57*. Cambridge University Press.

Serfling, R. J. (1980). Approximation theorems of mathematical statistics.

Steiner, S. H., Cook, R. J., Farewell, V. T., and Treasure, T. (200O). Monitoring surgical performance using risk-adjusted cumulative sum chart. *Bio-statistics*, *1(4):441-452*.

Wetterslev, J., Thorlund, K., Brok, J, and Gluud, C. (2008). Trial sequential analysis may established when firm evidence is reached in cumulative meta-analysis. *Journal of Clinical Epidemiology, 61(1):64-75*.