Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 4, Issue 1 (2011): 107-117

TESTIMATION OF AVERAGE LIFE IN PRESENCE OF SUSPECTED OUTLIERS IN EXPONENTIAL LIFE MODEL UNDER ASYMMETRIC LOSS FUNCTION

***Rakesh Srivastava**
Department of Statistics
Faculty of science
The M. S. University of Baroda
VADODARA - 390 002
*rakeshsrivastava30@yahoo.co.in

Vilpa Tanna
P. & S. M. Department
P. D. U. Medical College
Rajkot - 360 001.

Abstract

The present paper proposes a preliminary test estimator of average life (scale parameter) in two parameter exponential distribution in presence of suspected outliers. The risk properties of this testimator have been studied under asymmetric loss function and it is claimed that the proposed testimator dominates the never pool estimator (in terms of having smaller risk) in the whole range of life ratio considered here.

Key Words: Exponential model, scale parameter, outliers, asymmetric loss function, risk, life Ratio, risk ratio.

1. Introduction

An applied statistician who has analyzed a number of sets of real data is likely to have come across ‘outliers’. An outlier would be an observation (or subset of observations) which deviates so much from other observations as to arouse suspicions that it was generated by different mechanism. An inspection of a sample containing outliers would show up such characteristics as large gaps between outlying and in lying observations and the deviation between the outliers and the group of inliers as measured on some suitable standardized scale.

The effect of outliers on the analysis of a set of data depends strongly on the mechanism by which the outliers are believed to be generated. Usually, the major objective of the analysis will be to estimate a parameter of the distribution under study.

Let us suppose that out of n observations in the sample, $r(\geq 1)$ largest observations are suspected to be outliers. In other words, we have two samples of (n-r) and r observations respectively from two exponential populations having means θ_1 and θ_2.

The choice of r is subjective and requires judgement on the part of the experimenter. The choice of r is crucial for correctly rejecting observations as outliers. This choice, however, cannot be made before the experiment is done because no outliers are anticipated. Therefore, in practice the choice of r will have to be made subjectively after the experiment is done and data taken. In case this subjective choice of r is doubtful then follow the procedure given by Tietjen and Moore (1972). Also Daniel (1960) gives a probability plotting method for normal case. Kale (1974 b) gives a similar method for one parameter exponential density. Dempster and Rosner (1971) give some computational techniques based on specific prior distribution.

For testing the largest (smallest) observation as outlier in a sample from a normal population $N(\mu,\sigma^2)$, the most prominent statistic is due to Grubbs (1950). This statistic is the ratio of the sum of squares of deviations from sample mean for a reduced sample (obtained by omitting the largest (smallest) observation) to the sum of the squares of deviations for the whole sample. Its generalization for testing r largest (smallest) observations for outliers is due to Grubbs (1950, 1969) for $r = 2$ and Tietjen and Moore (1972) for general r . Kale (1979) has given a selective review of theory and methodology of statistical problems for data involving outlier or spurious observations.

The estimator of θ_1 will be a function of $n-r$ or n observations depending whether the r largest observations are outliers or not. For this purpose we first test the hypothesis that $\theta_1 = \theta_2$. The testing of such hypothesis before final inference is known as preliminary test of significance (PTS). The estimation of average life in two-parameter exponential population with type II censored data for two-sample problem using a PTS has been studied by Gupta and Singh (1985). For a detailed bibliography on PTS, see Han and Bancroft (1977), Han et. al. (1988).

For various life testing models, it has been observed that in the estimation of mean life or reliability function the use of squared error loss function (SELF) may be inappropriate as has been pointed out by Canfield (1970), Varian (1975), Zellner (1986), Basu and Ebrahimi (1991), Pandey and Rai (1992), Rai (1996) and Srivastava (1996). The above mentioned authors and many others have suggested to use asymmetric loss function for estimating or testimating the parameters or parametric functions in the context of life testing models. Varian (1975) proposed an asymmetric loss function, which has been found to be appropriate in the situations where either overestimation is more serious than the underestimation or vice-versa. The loss function is

$$L\left(\Delta\right) = b\left(e^{a\Delta} - a\Delta - 1\right), \qquad a \neq 0, \quad b > 0 \tag{1.1}$$

Where, $\Delta = (\hat{\theta} - \theta)$ or $\Delta = \left(\dfrac{\hat{\theta}}{\theta} - 1\right)$ depending upon the parameter (i.e. location or scale) which is being estimated.

LINEX loss function has two constants ‘*a*’ and ‘*b*’ which gives the freedom to tailor theloss according to our needs by choosing them appropriately.

The sign and magnitude of ‘*a*’ represents the direction and degree of asymmetry respectively. Positive values of ‘*a*’ are used when overestimation is more serious than underestimation, while negative values of ‘*a*’ are used in reverse situations. L(Δ) rises exponentially when $\Delta < 0$ and almost linearly when $\Delta > 0$. The loss function defined by (1.1) is known as the LINEX loss function. The factor of proportionality is ‘*b*’. It is ‘*a*’ which determines the relative losses for positive and negative values of estimation error. However for '*a*' close to zero this loss function is approximately 'SELF' and hence, almost symmetric. We have taken $b=1$ throughout our study.

The object of present investigation is to study the risk properties of the estimator obtained after a preliminary test of significance for the hypothesis $\theta_1=\theta_2$ in two parameter exponential distribution in presence of suspected outliers using asymmetric loss function.

In section 2 we have given the formulation of the problem, we have derived the risk of this estimator under asymmetric loss function in section 3. Section 4 comprises of numerical computations and recommendations regarding the application of proposed estimator.

2. Formulation of the Problem

Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample of size n drawn from a two-parameter exponential population:

$$f(X : A, \theta_1) = \theta_1^{-1} \exp\left\{-\frac{(X-A)}{\theta_1}\right\} \text{ for } X \geq A \text{ and } \theta_1 > 0$$

$$= 0, \quad \text{otherwise} \qquad (2.1)$$

where A and θ_1 are location and scale parameters respectively, both unknown. Let

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)} \qquad (2.2)$$

be the ordered sample by arranging $X_1, X_2, \ldots\ldots, X_n$ in ascending order of magnitude. A situation often encountered in practice is that r largest observations in (2.2) are suspected as outliers.

Suppose that the r largest observations in (2.2) constitute a random sample from $f(X; A, \theta_2)$ and the rest i.e. first (n-r) observations from $f(X; A, \theta_1)$. Now, we wish to estimate θ_1 but there is some uncertainty whether $\theta_1=\theta_2$. This uncertainty can be resolved by first testing the hypothesis H_0: $\theta_1=\theta_2$ against the alternative H_0: $\theta_1<\theta_2$ by using the test statistic proposed by Tiku (1975). The estimator thus obtained is called 'preliminary ***test estimator'***. Let $\hat{e}$ be the preliminary test estimator of θ_1. We propose the following point estimation procedure for the estimation of average life θ_1 as:

$$\hat{e} = \begin{cases} \dfrac{W}{(n-r-1)} & if \quad \dfrac{W}{(W+V)} \leq \beta \\[2ex] \dfrac{(W+V)}{(n-1)} & if \quad \dfrac{W}{(W+V)} > \beta \end{cases} \qquad (2.3)$$

where

$$W=\sum_{i=1}^{n-r}(X_{(i)}-X_{(1)})+r(X_{(n-r)}-X_{(1)}) , \qquad V=\sum_{i=n-r+1}^{n}(X_{(i)}-X_{(n-r)})$$

and β denotes the lower 100α% point of beta distribution with parameters (n-r-1,r)..

3. Derivation Of Risk of $\hat{e}$

Using the fact that $\frac{2W}{\theta_1}$ and $\frac{2V}{\theta_2}$ are independently distributed as central chi-square with 2(n-r-1) and 2r degrees of freedom respectively. The joint p. d. f. of W and V is given by

$$f_1(W,V)=k_0W^{n-r-2}V^{r-1}\exp\left\{-\left(\frac{W}{\theta_1}+\frac{V}{\theta_2}\right)\right\}dWdV \tag{3.1}$$

Where

$$k_0=\frac{1}{\left[\Gamma(n-r-1)\Gamma r\theta_1^{n-r-1}\theta_2^{r}\right]}$$

The risk of $\hat{e}$ under L(Δ) is given by R ($\hat{e}$ **),**
Where

$$R(\hat{e})=E\left[L\left(\hat{e},\hat{\theta}_1\right)\right]$$

or,

R($\hat{e}$ **) =** E$_1$P$_1$ + E$_2$P$_2$ (3.2)

where

$$E_1=E\left[L\left(\frac{W}{n-r-1}\Big\backslash\frac{W}{(W+V)}\le\beta\right)\right],$$

$$E_2=E\left[L\left(\frac{W+V}{n-r}\Big\backslash\frac{W}{(W+V)}>\beta\right)\right]$$

and

$$P_1=P\left(\frac{W}{(W+V)}\le\beta\right), \quad P_2=P\left(\frac{W}{(W+V)}>\beta\right)$$

To calculate the above expectations let us make the following transformations in (3.1):

$$t_1=W/(W+V), \quad t_2=W+V$$

On simplification, we get

$$f_2(t_1,t_2) = k_0 t_1^{n-r-2}(1-t_1)^{r-1} t_2^{n-2} \exp\left[\frac{-t_2\left\{\phi+\left(1-\phi\right)t_1\right\}}{\theta_1}\right] dt_1 dt_2 \tag{3.3}$$

Where

$\phi = \dfrac{\theta_1}{\theta_2}$ With $\phi < 1$, which represents the life ratio

Now,

$$\begin{aligned} R\left(\hat{e}\right) &= \int_{t_1=0}^{\beta}\int_{t_2=0}^{\infty}\left[e^{a\left\{\frac{W}{(n-r-1)\theta_1}-1\right\}} - a\left\{\frac{W}{\left(n-r-1\right)\theta_1}-1\right\}-1\right] f_2\left(t_1,t_2\right) dt_1 dt_2 \\ &\quad + \int_{t_1=0}^{\beta}\int_{t_2=0}^{\infty}\left[e^{a\left\{\frac{W+V}{(n-r)\theta_1}-1\right\}} - a\left\{\frac{W+V}{\left(n-r\right)\theta_1}-1\right\}-1\right] f_2\left(t_1,t_2\right) dt_1 dt_2 \end{aligned} \tag{3.4}$$

A straightforward integration of (3.4) leads us to

$$\begin{aligned} R\left(\hat{e}\right) = \Bigg[& \frac{e^{-a} I\left(1-X_1;\, n-r-1, r\right)}{\left(1-\dfrac{a}{n-r-1}\right)^{n-r-1}} + \frac{e^{-a} I\left(X_2;\, r, n-r-1\right)}{\left(1-\dfrac{a}{n-1}\right)^{n-r-1}\left(1-\dfrac{a}{\phi\left(n-1\right)}\right)^{r}} \\ & - \frac{ar}{n-1}\left\{\frac{I\left\{X_0;\, r+1, n-r-1\right\}}{\phi} - I\left\{X_0;\, r, n-r\right\}\right\} - 1 \Bigg] \end{aligned} \tag{3.5}$$

where

$$B\left(X;\, m,n\right) = \frac{\int_{x}^{1} y^{m-1}(1-y)^{n-1} dy}{B(m,n)} = 1 - I\left(X;\, m,n\right) = I\left(1-X;\, n,m\right)$$

with

$$I\left(X;\, m,n\right) = \frac{\int_{0}^{x} y^{m-1}(1-y)^{n-1} dy}{B(m,n)} \quad \text{and} \quad X_0 = \frac{\phi\left(1-\beta\right)}{\beta+\phi\left(1-\beta\right)},$$

$$X_1 = \frac{\phi(1-\beta)}{\left(1-\frac{a}{n-r-1}\right)\beta + \phi(1-\beta)}, \; X_2 = \frac{\left(\phi - \frac{a}{n-1}\right)(1-\beta)}{\left(1-\frac{a}{n-1}\right)\beta + \left(\phi - \frac{a}{n-1}\right)(1-\beta)}$$

4. Risk Comparison

A natural way of comparing the behaviour of the proposed estimator is to examine the performance of it with respect to the best available estimator, for this purpose we define the relative risk of $\hat{e}$ with respect to never pool estimator is given by

$$R_R = \frac{R_N}{R(\hat{e})} \tag{4.1}$$

where R_N is the risk of the estimator never pool estimator under L(Δ).
Now,

$$R_N = \int_{t_1=0}^{\infty}\int_{t_2=0}^{\infty}\left[e^{a\left\{\frac{W}{r\theta_1}-1\right\}} - a\left\{\frac{W}{r\theta_1}-1\right\} - 1\right] f_2(t_1,t_2)\,dt_1 dt_2 \tag{4.2}$$

A straightforward integration of (4.2) gives

$$R_N = \frac{e^{-a}}{\left(1-\frac{a}{r}\right)^r} - 1 \tag{4.3}$$

Hence, R_R=[expression (4.3)][expression (3.5)]$^{-1}$

Evidently R_R is a function of n. r, n, r, ϕ, α and '*a*'. We have considered some values of ϕ i.e. ϕ=0.1(0.1)1.0. We have taken two data sets for n and r i.e. n=30, r=10 and n=50, r=17. The values of α's are taken to be 1%, 5%, 10% and 25%. We have selected positive as well as negative values of '*a*' to observe the effect of overestimation or vice-versa i.e. $a = \pm 1, \pm 2, \pm 3, \pm 4, \pm 5$. The graphs of R_R for all the above mentioned values have assembled at the end of the paper.

It is observed that $\hat{e}$ performs better than the usual estimator for '*a*' ranging from 1 to 5, $0.1 \le \phi \le 1.0$. At 25% level of significance the magnitude of R_R is highest though it is good for all the other values of α's also considered here, including lower values such as 1% and 5%.

It seems that due to presence of outliers a larger gain in R_R is attained at higher α and for fairly large degree of asymmetry. But studying the risk properties of the preliminary test estimator under asymmetric loss function has definitely an edge over that studied under squared error loss function. Proper choices of 'a' guide the experimenter so as how to handle the situation with outliers. For negative values of '*a*' it is observed that the best performance is attained for first data set for α=25% and for $a=-1\ \ to\ \ -5,\ 0.1 \le \phi \le 1.0$. Here again a lower level of significance can be considered say α=1% or 5% as the performance of $\hat{e}$ is better for these values too. However comparing the performances for both the sets, it is evident that for the second data set i.e. n=50, r=17 there is higher gain for lower '*a*'. However when $a \ge 3$ the higher values of R_R are obtained for the first data set.

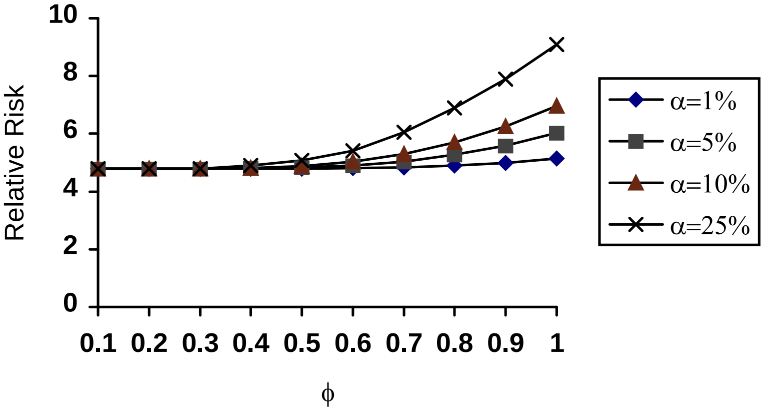
References

1. Bancroft, T. A. and Han, C. P. (1977). Inference based on conditional specification: A note and a bibliography, Int. Stat. Rev., 45, p.117-127.
2. Basu, A. P. and Ebrahimi, N. (1991). Bayesian approach to life testing and reliability estimation using asymmetric loss function, JSPI, 29, p. 21-31.
3. Canfield, R.V. (1970). A Bayesian approach to reliability estimation using a loss function, IEEE transformations on Reliab., 19, p. 13-16.
4. Daniel, C. (1960). Locating outliers in factorial experiments, Technometrics, 2, p. 149-156.
5. Dempster, A. P. and Rosner, B. (1971). Detection of outliers, Statistical decision theory and related topics, p. 161-180 (Ed. S. S. Gupta), Academic Press.
6. Grubbs, F. E. (1950). Sample criteria for testing outlying observations, Ann. Math. Statist., 21, p. 27-58.
7. Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples, Technometrics, 11, p. 1-21.
8. Gupta, V. P. and Singh, Umesh (1985). Preliminary test estimator for life data, Microelect Reliab., 25(5), p. 881-887.
9. Han, C. P., Rao, C. V. and Ravichandran, J. (1988). Inference based on conditional specification: A second bibliography, Comm. Stat. -TM, 17(6), p. 1945-1966.
10. Kale, B. K. (1974 b). Detection of outlier, (Tech. Report 63, Deptt. of Statistics, University of Manitoba) paper presented at the international symposium on recent trends of research in statistics, held at I. S. I., Calcutta, December, 1974, appeared in $Sankhya\ \bar{a}$, 38(B), p. 356-363.
11. Kale, B. K. (1979). Outliers - A Review, Jr. Ind. Statist. Assoc., 17, p. 51-67.
12. Pandey, B. N. and Rai, O. (1992). Bayesian estimation of mean and square of mean of normal distribution using LINEX loss function, Comm. in Stat. Theory and Method, 21 (2), p. 3386-3391.
13. Rai, O. (1996). A some times pool estimator of mean life under LINEX loss function, Comm. in Stat. Theory and Method, 25, p. 2057-2067.
14. Srivastava, R. (1996). Bayesian estimation of scale parameter and reliability in weibull distribution using asymmetric loss function, IAPQR transactions, 21 (2), p. 143-148 .
15. Srivastava,R. And Dulawat, M. S. (2002). Preliminary test estimator for average life of Exponential distribution in presence of suspected outliers, IAPQR transactions, Vol.27, No.1.
16. Tietjen, G. L. and Moore, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers, Technometrics, 14, p. 583-597.

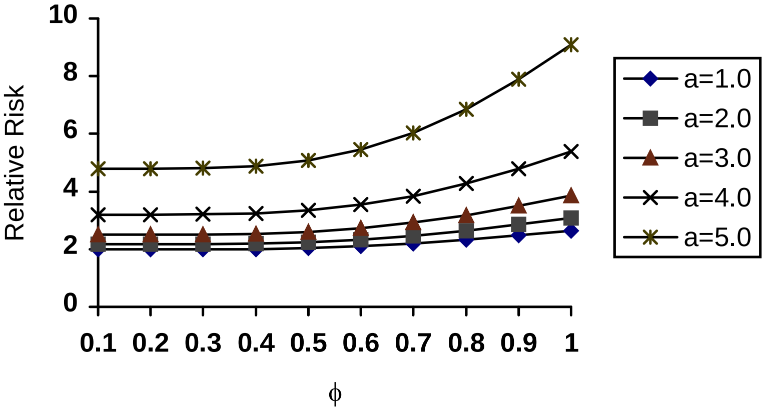
17. Tiku, M. L. (1975). A new statistic for testing suspected outliers, Comm. Stat. 4(8), p. 737-752.
18. Varian, H. R. (1975). In S. E. Fienberg and A. Zellner, ed. Studies in Bayesian econometrics and statistics in honour of Savage, North Holland, Amsterden, p. 195-208 .Zellner, A. (1986). Bayesian estimation and predictions using asymmetric loss function, JASA, 61, p. 446-451.

GRAPHS OF R_R PRELIMINARY TEST ESTIMATOR

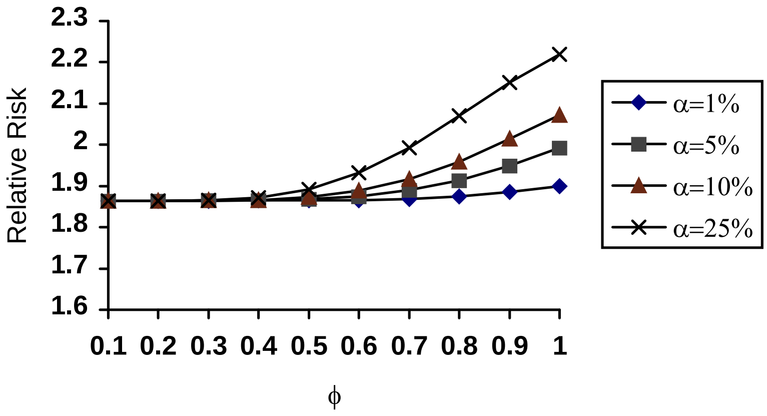
(1) $n = 30, \quad r = 10, \quad a = 5.0$



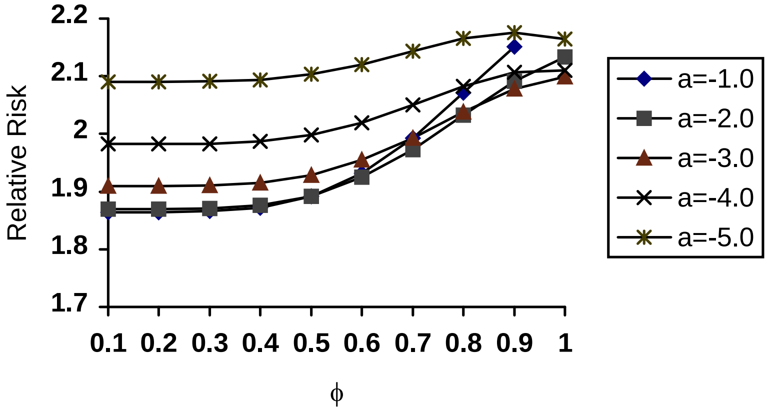
(2) $n = 30, \quad r = 10, \quad \alpha = 25\%$



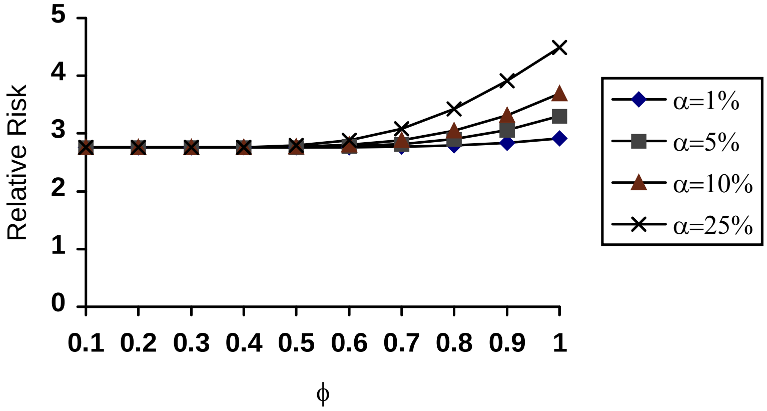
(3) $n = 30, \quad r = 10, \quad a = -5.0$



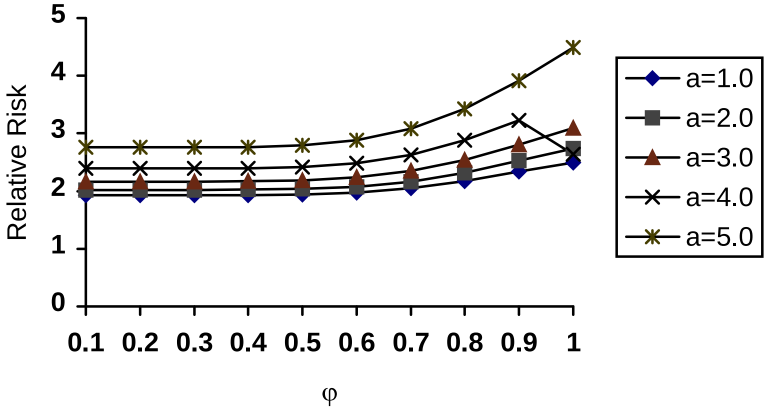
(4) $n = 30, \quad r = 10, \quad \alpha = 25\%$



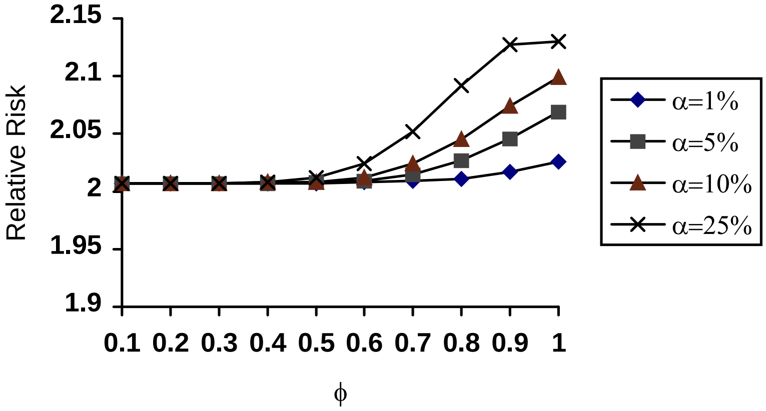
(5) $n = 50, \quad r = 17, \quad a = 5.0$



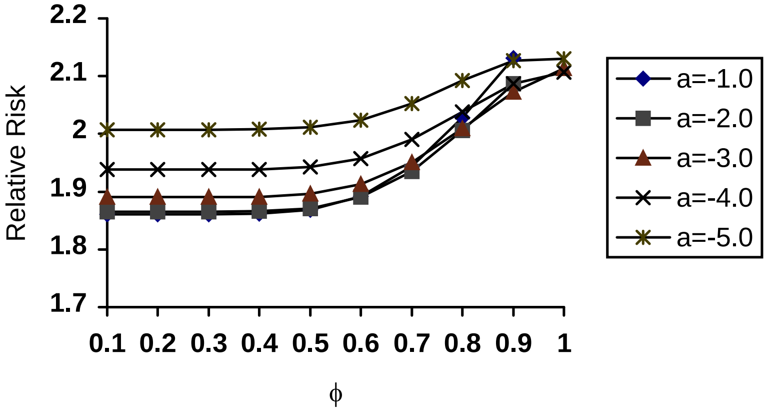
(6) $n = 50, \quad r = 17, \quad \alpha = 25\%$



(7) $n = 50, \quad r = 17, \quad a = -5.0$



(8) $n = 50, \quad r = 17, \quad \alpha = 25\%$


Relative Risk
ϕ
a=-1.0
a=-2.0
a=-3.0
a=-4.0
a=-5.0