Mathematical and Computational Applications, Vol. 18, No. 2, pp. 94-102, 2013

OPTIMAL RESTRICTED DESIGNS FOR THE INVERSE GAUSSIAN MEAN

Sevil Bacanlı

Hacettepe University, Faculty of Science, Department of Statistics, 06800 Beytepe Ankara-Turkey.
sevil@hacettepe.edu.tr

Abstract-This study deals with the development of optimal restricted two and three stage designs when response variable has an inverse Gaussian distribution with known scale parameter.

Key Words- Optimal restricted two and three-stage design, Inverse Gaussian distribution

1. INTRODUCTION

Fixed sample size design is not useful in experiment that subjects enter to the study sequentially. Consequently it is possible to analyze the accumulated data sequentially. Wald [1], introduced sequential analysis and demonstrating that sequential probability ratio test (SPRT) require substantially fewer observations than a fixed sample test of equivalent statistical power.

SPRT is widely used in clinical trials, quality control studies and life tests. Also generally sequential designs cannot be used in such situations. In this case analyzing the accumulated data in groups is the most convenient way.

Data are analyzed after groups of observations are entered into a group sequential design. However group sequential designs are generally more practical and they provide much of the saving possible from sequential designs [2].

Group sequential designs are widely used in clinical trials. In most randomized clinical trials with sequential patient entry, fixed sample size design is unjustified on ethical grounds and sequential designs are often impractical.

Two and three stage design is the simplest form of a group sequential design. Case *et al.* [3], [4], developed optimal restricted two (OR_2) and three (OR_3) stage design that have the restriction of using the fixed sample critical value at the final stage.

In general, optimal restricted two and three-stage designs has been proposed for normally and binomial response variable. In this study, a optimal restricted designs when response variable has an inverse Gaussian distribution with known scale parameter is proposed.

Inverse Gaussian distribution function is a very useful alternative to the real life time distribution such as gamma, log-normal and Weibull distributions. The distribution has a wide application area in clinical trials, quality and reliability theory, industrial engineering applications and life tests. In those areas, outcome variable can be measured in series because data is accumulated sequentially. [5, 6]

Edgeman and Salzberg [6] and Edgeman and Lin [7] developed the sequential probability ratio test for the inverse Gaussian mean, and its application to sequential sampling plans. Bacanlı and Demirhan [8] suggested the group sequential test when response variable has an inverse Gaussian distribution with known scale parameter.

This study is organized as follows: In section 2, the optimal restricted designs are described. In section 3, SPRT for the mean of inverse Gaussian distribution is briefly reviewed and it is shown that optimal restricted designs can be used in inverse Gaussian mean. Example and the optimal restricted designs comparison to other design are given in section 4.

2. OPTIMAL RESTRICTED DESIGNS

In this section**,** firstly OR_2 design is examined for response variable has normal distribution with mean (θ) and known variance (σ^2).
For testing $H_0:\theta=\theta_0$ against H_1: $\theta>\theta_0$, the OR_2 design is defined as follows;

Stage I: Accrue n_1 observations and calculate test statistic,

$$Z_1=\frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \tag{1}$$

where $\hat{\theta}$ is calculated from data on the first n_1 observation. If $Z_1<C_1$; Accept H_0, if $Z_1>C_2$; Reject H_0, otherwise; continue the second stage.

Stage 2: Accrue an additional n_2 observations. Let $n=n_1+n_2$ and calculate,

$$Z=\frac{\hat{\theta}-\theta_0}{\sigma_{\hat{\theta}}} \tag{2}$$

where $\hat{\theta}$ is computed from data on all n observations. If $Z<C_3$; Accept H_0, otherwise, reject H_0 [2].

Z_1 and Z are distributed standard normal distribution and their joint distribution is bivariate normal with zero means, unit variances, and correlation $(n_1/n)^{1/2}$ .The maximum sample size for the two-stage design is n and is realized whenever a second stage is necessary. The expected sample size (ESS) of the two-stage design is given by equation (3):

$$ESS(\theta)=n\ [1-(1-p)P_S(\theta)] \tag{3}$$

where $P_S(\theta)$ denote the probability that the trial will be stopped at the first stage, and p is the rate of the number of observations at the first stage to the number of total observations at the second stage $p=n_1/n$. For some studies it might be practical to choose equal samples at each stage. Therefore, if p=0.50, each stage have equal sizes. θ

value can be computed for θ_0, θ_1 where θ_0 is the θ value when H_0 is true; θ_1 is the θ value when H_1 is true.
There are five unknown parameters in the two-stage design, namely: n_1, n_2, C_1, C_2 and C_3. The critical value at the second stage, C_3, will be set to equal that of the fixed sample test

$$C_3 = \varphi^{-1}(1-\alpha) \quad (\text{or } \varphi^{-1}(1-\frac{\alpha}{2})) \tag{4}$$

where $\varphi(x)$ denotes the standard normal distribution function. The other four parameters of interest are chosen to satisfy the two equations:

$$\alpha = 1-\Phi(C_2) + B(C_1, C_2; C_3, \infty; p) \tag{5}$$

$$1-\beta = 1-\Phi(C_2 - u\sqrt{p}) + B(C_1 - u\sqrt{p}, C_2 - u\sqrt{p}; C_3 - u, \infty; p) \tag{6}$$

where,

$$B(a,b,c,d,p) = \left(1/2\pi\sqrt{1-p}\right)\int_a^b\int_c^d \exp\left\{-(1/2)(1-p)\left(y^2 - 2\sqrt{p}yz + z^2\right)\right\}dy\,dz \tag{7}$$

and $u = \sqrt{n}(\theta_1 - \theta_0)/\sigma$.

Equation (5) and (6) are solved iteratively by numerical integration of the bivariate normal distribution using a double precision function [3, 9].

With five parameters and only three constraints given by equations (4), (5), (6) optimality criteria are used to determine the parameter values. So, this test is called optimal restricted two-stage design. In this study, we have examined Bayes criteria.

<u>Bayes Criterion:</u>
Minimize a weighted average of the ESS under H_0 and the H_1,

$$\text{minimize } ESS_w(\theta) = (1-w)ESS(\theta_0) + wESS(\theta_1) \tag{8}$$

Using a weight of 0 for this criterion gives the most efficient designs if the null hypothesis is true while a weight of 1 gives the most efficient designs if the specified alternative is true .

The optimal design parameters, the probabilities $p_s(\theta)$, maximum (n) and expected sample sizes (ESS) obtained using the bayes criteria are listed in Table 1 for p=0.50 and several values of α and $1-\beta$. In tables, n_f is the sample size for a fixed sample design[3,9].

The sample size required for a OR_2 design is obtained by multiplying the tabled values by $\Delta^2 = \{\sigma/(\theta_1-\theta_0)\}^2$

Table 1. Optimal restricted two-stage one- sided designs for bayes criterion at given $\alpha = 0.01, 0.05$ $1-\beta = 0.80, 0.90$ p=0.50

w	α	1-β	p	C_1	C_2	C_3	n_f [a]	n [a]	ESS(θ_0) [a]	ESS(θ_1) [a]
0	0.01	0.80	0.5	1.052	2.833	2.326	10.036	10.849	6.212	8.641
		0.90	0.5	1.014	2.856	2.326	13.017	14.085	8.123	10.778
	0.05	0.80	0.5	0.638	2.150	1.645	6.183	6.907	4.303	5.194
		0.90	0.5	0.595	2.178	1.645	8.564	9.558	6.029	6.886
1	0.01	0.80	0.5	1.310	2.690	2.326	10.036	11.612	6.343	8.561
		0.90	0.5	1.253	2.720	2.326	13.017	15.009	8.266	10.687
	0.05	0.80	0.5	0.768	2.066	1.645	6.183	7.203	4.328	5.175
		0.90	0.5	0.700	2.109	1.645	8.564	9.874	6.046	6.864

[a] Multiply each value by $\{\sigma/(\theta1-\theta0)\}^2$

OR_3 design is an extension of the OR_2 design to three stages. However the sample sizes must be equal for each stage of the design [4].

The OR_3 design for normal mean testing is given as follows:

Stage I: Accure n_1 observations and calculate test statistics ,

$$Z_1 = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \tag{9}$$

Where $\hat{\theta}$ is calculated from data on the first n_1 observation. If $Z_1 < C_1$; Accept H_0, if $Z_1 > C_2$; Reject H_0 , otherwise; continue the second stage.

Stage 2: Accrue an additional n_2 observation. Let $n = n_1 + n_2$ and calculate,

$$Z_2 = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \tag{10}$$

where $\hat{\theta}$ is computed from data on all n observations. If $Z_2 < C_3$; Accept H_0, if $Z_2 > C_4$; Reject H_0 , otherwise; continue the second stage.

Stage 3: Accrue an additional n_3 observation. Let $n = n_1 + n_2 + n_3$ and calculate,

$$Z_3 = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \tag{11}$$

where $\hat{\theta}$ is computed from data on all n observations. If $Z_3 < C_5$; Accept H_0, otherwise, reject H_0.

There are eight unknown parameters in the OR_3 design, namely n_1, n_2, n_3,C_1, C_2, C_3, C_4, and C_5. The critical value at the final stage, will be set equal to that of the fixed sample test.

OR_3 design considers the case of equal sample sizes at each stage, reducing the number of unknown parameters to six.

With six parameters and only two constraints, parameter values are chosen to min ESS(θ) for Bayes criteria. Therefore the algorithm used to obtain the parameter values for OR_3 design is olmost identical in the OR_2 [4, 9].

The design parameters and the sample sizes obtained using Bayes criteria for OR_3 design are given in Table 2 for $\alpha = 0.01, 0.05$ $1-\beta = 0.80, 0.90$.

Table 2. Optimal restricted three-stage one- sided designs for bayes criterion at given $\alpha = 0.01, 0.05$ $1-\beta = 0.80, 0.90$

w	α	$1-\beta$	C_1	C_2	C_3	C_4	C_5	n_f [a]	n [a]	$ESS(\theta_0)$ [a]	$ESS(\theta_1)$ [a]
0	0.01	0.80	0.738	3.819	1.338	2.598	2.326	10.036	11.642	5.018	8.430
		0.90	0.649	3.747	1.335	2.632	2.326	13.017	15.100	6.639	10.544
	0.05	0.80	0.342	2.539	0.877	1.945	1.645	6.183	7.543	3.710	4.946
		0.90	0.234	2.470	0.879	2.015	1.645	8.564	10.362	5.310	6.423
1	0.01	0.80	0.816	2.796	1.724	2.661	2.326	10.036	12.646	5.219	8.029
		0.90	0.535	2.719	1.907	2.696	2.326	13.017	16.662	7.290	9.763
	0.05	0.80	0.312	2.150	1.184	2.023	1.645	6.183	7.976	3.833	4.823
		0.90	0.012	2.095	1.313	2.067	1.645	8.564	11.048	5.738	6.252

[a] Multiply each value by $\{\sigma/(\theta1-\theta0)\}^2$

3. OPTIMAL RESTRICTED DESIGNS FOR THE MEAN OF AN INVERSE GAUSSIAN DISTRIBUTION

Let x is an inverse Gaussian (IG) distributed random variable and its probability density function is defined as follows;

$$f(x,\mu,\lambda) = \left[\frac{\lambda}{(2\pi x^3)}\right]^{1/2} \exp\left[\frac{-\lambda(x-\mu)^2}{(2x\mu^2)}\right]; \qquad x > 0, \mu > 0, \lambda > 0 \tag{12}$$

Here μ is the mean of the distribution. So it is a location parameter and λ is a scale parameter [5].

Given a sequence of observations $x_1, x_2, \ldots$ from inverse Gaussian distribution (12), suppose one wishes to test the simple null hypothesis H_0:$\mu=\mu_0$ against the simple alternative H_1: $\mu=\mu_1$ ($\mu_1 < \mu_0$) , when λ is known. The SPRT for testing H_0 is defined as follows:

Let

$$\sum_{i=1}^{n} z_i = \frac{\lambda}{2}\left[\sum_{i=1}^{n} x_i\left[(\mu_1^2-\mu_0^2)/(\mu_0\mu_1)^2\right] - \left[2n(\mu_1-\mu_0)/(\mu_0\mu_1)\right]\right] \tag{13}$$

At the nth stage, accept H_0 if $\sum_{i=1}^{n} z_i \le \ln B$, reject H_0 If $\sum_{i=1}^{n} z_i \ge \ln A$, otherwise, $\ln B < \sum_{i=1}^{n} z_i < \ln A$ continue sampling by taking an additional observation. If α and β

are the type I and type II errors respectively, then according to SPRT, A and B are approximately given by $A \approx (1-\beta)/\alpha$ and $B \approx \beta/(1-\alpha)$ [1, 6].

The average sample number (ASN) function under H_0 and H_1 is approximately given (14) and (15),

$$E(n;\mu_0) \approx \frac{(1-\alpha)\ln B + \alpha \ln A}{E(Z;\mu_0)} \tag{14}$$

$$E(n;\mu_1) \approx \frac{\beta \ln B + (1-\beta)\ln A}{E(z;\mu_1)} \tag{15}$$

$$E(z;\mu) = \frac{\lambda}{2}\left[\mu\left(\mu_1^2 - \mu_0^2\right)/(\mu_0\mu_1)^2 - 2(\mu_1 - \mu_0)/(\mu_0\mu_1)\right].$$

IG distribution is related to normal distribution. This relation was given with following theorem in Chikara and Folks [5], which establishes a basic relationship between IG and the normal.

Theorem: Let $Y = \sqrt{\lambda}(X-\mu)/\mu\sqrt{X}$. Then the pdf of Y is given by

$$f(y) = \left(1 - \frac{y}{\sqrt{4\lambda/\mu + y^2}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\right), \quad -\infty < y < \infty \tag{16}$$

The transformation $Y = \sqrt{\lambda}(X-\mu)/\mu\sqrt{X}$ is one-to-one and as x varies from 0 to ∞, y varies from $-\infty$ to ∞.

Then the cumulative distribution function $F(y)$ of Y is

$$F(y) = \Phi(y) + e^{2\lambda/\mu}\Phi\left(-\sqrt{4\lambda/\mu + y^2}\right), \qquad -\infty < y < \infty$$

where Φ is the standard normal distribution function. In this case $F(y) \to \Phi(y)$ as $\Phi = \lambda/\mu \to \infty$. Because of this and because of the one-to-one relationship between x and y, one finds that the distribution of X is asymptotically normal with mean μ and variance μ^3/λ [5].

Therefore fixed sample test for the mean of an inverse Gaussian distribution based on the standard normal distribution.

Given a random sample $x_1, x_2, \ldots, x_n$ drawn from IG distribution. Consider the testing of the hypothesis, $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, when λ is known, the test statistic for the mean is defined as,

$$Z = \frac{\sqrt{n\lambda}(\bar{x} - \mu_0)}{\mu_0 \sqrt{\bar{x}}} \tag{17}$$

Here, $\bar{x} \sim IG(\mu_0, n\lambda)$. The test statistic is compared with $Z_{1-\alpha/2} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ where ϕ denotes the standard normal distribution function. Consequently if $|Z| > Z_{1-\alpha/2}$ then H_0 is rejected for two sided hypothesis [5, 8].

In the sense of this information, we modify restricted optimal two-stage design for the mean of an inverse Gaussian distribution. Test statistics of OR_2 can be defined from equation (1), (2), and (17),

Stage I: Accrue n_1 observations and calculate test statistic,

$$Z_1 = \frac{\sqrt{n\lambda}(\bar{x}_1 - \mu_0)}{\mu_0 \sqrt{\bar{x}_1}} \tag{18}$$

Where $\bar{x}$ is calculated from data on the first n_1 observation.

Stage II: Accrue additional n_2 observations. Let $n = n_1 + n_2$ and calculate,

$$Z = \frac{\sqrt{n\lambda}(\bar{x} - \mu_0)}{\mu_0 \sqrt{\bar{x}}} \tag{19}$$

Where $\bar{x}$ is calculated from data on the first n_1 observation.

According to Theorem 1, Z_1 and Z have a standard normal distribution. Therefore, it is suggested that the OR_2 and OR_3 designs can be used for testing inverse Gaussian distribution mean with known scale parameter.

OR_3 design is an extension of OR_2 design to three stages and test statistics Z_i can be obtained as given OR_2 design.

In this case, design parameters for OR_2 and OR_3 designs are the same as normal distribution parameters. However, the sample size is obtained by multiplying the values (Table1-2) by $\Delta^2 = \left(\frac{\mu_0^2 \mu_1}{\lambda(\mu_1 - \mu_0)^2}\right)$. Therefore, the only change in design is the sample size.

4. COMPARISON WITH OTHER DESIGN AND DISCUSSION

In this section, the comparison of optimal restricted designs (OR_2 and OR_3) with fixed sample design and SPRT for inverse Gaussian distribution have been examined with an example.

As an example, suppose $H_0 : \mu = 0{,}03$ against $H_1 : \mu = 0{,}05$ and $\lambda = 0{,}1$ α=0.05 1-β=0.90 Results are given in Table 3.

Table 3. Comparison of fixed, SPRT and OR_2, OR_3 designs

Design	n	ESS(H_0)	ESS(H_1)	R(μ_1)
Fixed sample	9.634	9.634	9.634	-
SPRT	∞	7.667	5.346	-
OR_2				
W=0	10.753	6.783	7.747	0.44
W=1	11.108	6.802	7.722	0.46
OR_3				
W=0	11.657	5.974	7.226	0.56
W=1	12.429	6.455	7.034	0.60

It is well known that the SPRT has the minimum ESS(H_1), but it hasn't got a finite maximum number of observation. Furthermore OR_3 design needs a smaller sample size than other designs. In this study, ıt is seen that these results are also valid for inverse Gaussian distribution [1,4].

Let $S_{SPRT}(\mu_1) = n_f - ESS_{SPRT}(\mu_1)$ i=0,1 denote the savings possible with the use of the SPRT design. Also $S_2(\mu_1) = n_f - ESS_2(\mu_1)$ and $S_3(\mu_1) = n_f - ESS_3(\mu_1)$ denote the savings possible with optimal restricted two and three stage designs. Then the ratios $R_2 = S_2(\mu_1)/S_{SPRT}(\mu_1)$ and $R_3 = S_3(\mu_1)/S_{SPRT}(\mu_1)$ give the proportion of the possible savings realized with two and three–stage designs. A comparison of sequential design and the optimal restricted design according to proportion of the possible savings is given Table 3.

In Table 3, it is clear that OR_2 design can provide approximately 50% of the savings that would have been realized with SPRT design. OR_3 design provides as much as 60% of the possible savings.

Other examples with different choices of α, β, μ, and λ could readily be presented to illustrate the same general principle.

Case at all [4], compared those designs for the mean of normal distribution and obtained similar results. Therefore, as for normal distribution, it is advantageous to use optimal restricted two and three-stage design for the mean of inverse Gaussian distribution.

5. REFERENCES

1. A. Wald, A., Sequential Analysis, N.Y., John Wiley , New York, 1947.
2. C. Jennison, B.W Turnbull, *Group Sequential Methods with Application Clinical Trials*, Chapman and Hall, New York, 2000.
3. L.D. Case, T.M. Morgan, C.E. Davis, Optimal restricted two-stage designs, *Controlled Clinical Trials* **8**, 146-156, 1987.
4. L.D. Case, T.M. Morgan, C.E. Davis, Three-stage designs for monitoring clinical trials, *Communications in Statistics-Theory and Methods* **23 (7)**, 1875-1893, 1994.
5. R.S.Chhikara, J.L Folks, *The Inverse Gausssian Distribution:Theory, Methodology, and Applications*, Marcel Dekker, New York, 1988.
6. R.L.Edgeman, P.M Salzberg, Sequential sampling plan for inverse gaussian mean, *Statistical Papers* **32**, 45-53, 1991.
7. R.L. Edgeman, D.K.J Lin, Sequential analysis of an accelerated life model, *International Journal Quality and Reliability Management* **14 (6),** 598-605, 1997.
8. S. Bacanlı, Y. Parlak Demirhan, A Group Sequential Test for the Inverse Gaussian Mean , *Statistical Papers* **49**, 377-386, 2008.
9. S. Bacanlı, Optimal three-stage design, *Hacettepe Journal of Mathematics and Statistic* **35(2**), 227-234, 2006.