BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 15, Number 1, March 1987

ON THE CHOICE OF THE SMALLEST SAMPLE SIZE FOR SOME POPULATIONS

BY

C. C. SONG (宋傳欽)

Abstract. The property of a strictly decreasing sequence is applied to find the smallest sample size for the population with an exponential or a Poisson distribution when the upper bounds for the probabilities of two type errors are preassigned. A lower bound for the smallest sample size is also derived for each one of the two populations.

1. **Introduction.** Throughout this paper we are interested in testing a simple null hypothesis $H_0 : \theta = \theta_0$ against a simple alternative hypothesis $H_a : \theta = \theta_1$, where θ is the parameter of the distributon under study, and $\theta_0 > \theta_1$ is assumed. It is well-known that, given the upper bounds of the probabilities of two type errors, the sequential probability ratio test procedure is superior to the fixed sample size test procedure, since, on the average, the former saves more observations. In most books, e. g. Rao [1] and Rohatgi [2], the only example provided is the population with normal distribution when these two test procedurces are compared. If the problems of finding the smallest sample size for other distributions can be solved, then more examples can be provided for the comparison of these two test procedures.

Suppose $f_n(x) = x^{n-1} e^{-x} / (n-1)!$, $x > 0$, $n = 1, 2, \cdots$, and α, β with $0 < \alpha$, $\beta < 1$, $\alpha + \beta < 1$ are given.

Let a_n, b_n satisfy $\int_0^{a_n} f_n(x)\,dx = \beta$ and $\int_{b_n}^{\infty} f_n(x)\,dx = \alpha$.

Received by the editors August 22, 1986 and in revised form November 24, 1986.
Keywords: Exponential distribution, Poisson distribution; Gamma distribution; Stirling's formula.

First, we show a useful lemma that $\{b_n/a_n\}$ is a strictly decreasing sequence. Then, this property is applied to solve the following problems: Given α and β, choose the smallest sample size for the population with an exponential or a Poisson distribution so that the probabilities of two type errors do not exceed α and β.

For the population with an exponential distribution, it is in general unlikely to obtain a most powerful test with probabilities of two type errors exactly equal to α and β respectively, i. e., there usually exists no sample size n such that

$$(1) \qquad P_r(\bar{x} > k_n | \theta = \theta_0) = \alpha, \qquad P(\bar{x} < k_n | \theta = \theta_1) = \beta,$$

unless the ratio θ_0/θ_1 satisfies certain condition. When this fails, then we search for the samllest sample size n with

$$(2) \qquad P_r(\bar{x} > k_n | \theta = \theta_0) \leq \alpha, \qquad P_r(\bar{x} < k_n | \theta = \theta_1) \leq \beta.$$

Next, consider the population with a Poisson distribution. Just as we have mentioned for the exponential distribution case, only under a very restrictive condition, it is then possible to find the smallest sample size n satisfying

$$(3) \qquad P_r\left(\bar{x} < \frac{k_n}{n} \middle| \theta = \theta_0\right) = \alpha, \qquad P_r\left(\bar{x} \geq \frac{k_n}{n} \middle| \theta = \theta_1\right) = \beta,$$

in which k_n is a positive integer to be determined. Otherwise, we choose the smallest sample size n with

$$(4) \qquad P_r\left(\bar{x} < \frac{k_n}{n} \middle| \theta = \theta_0\right) \leq \alpha, \qquad P_r\left(\bar{x} \geq \frac{k_n}{n} \middle| \theta = \theta_1\right) \leq \beta.$$

Notice that, in (1) and (3), α and β are then the probabilities of two type errors; yet, in (2) and (4), α and β are interpreted as the pre assigned upper bounds for the probabilities of two type errors. Although, for convinence, we use the same notation k_n in (1), (2), (3) and (4), once we mention k_n, the readers can easily distinguish which k_n we are talking about.

By using the Stirling's formula, a lower bound for the smallest sample size is also derived for each one of the two populations.

At the end of this paper we illustrate how to make use of the chi-square table to obtain the smallest sample size for each population, and the strictly decreasing property of the sequnce $\{b_n/a_n\}$ can help us save time for doing that.

2. Statement and proof of the lemma. In this section we prove a useful lemma which will be used latter.

Consider the following two equations:

$$\sum_{t=0}^{n-1} \frac{x^t e^{-x}}{t!} = 1-\beta, \tag{5}$$

$$\sum_{t=0}^{n-1} \frac{x^t e^{-x}}{t!} = \alpha. \tag{6}$$

It can be shown that, for each n, a_n and b_n, defined as in Section 1, are respective roots of (5) and (6).

LEMMA. *$\{b_n/a_n\}$ is a strictly decreasing sequence.*

Proof. Fix n, and let $b_{n+1}=b_n+\delta$, $a_{n+1}=a_n+\theta, b_n/a_n=c$. Note that the θ appears in the proof of this lemma is not the parameter under study. Since $1-\alpha>\beta$, it follows that $c>1$. Now we claim that $\delta<c\theta$. From (5), we have

$$\begin{aligned}\sum_{t=0}^{n} \frac{(a_n+\theta)^t}{t!} &= (1-\beta)\, e^{a_n+\theta}\\ &= (1-\beta)e^{a_n} e^{\theta}\\ &= \sum_{t=0}^{n-1} \frac{a_n^t}{t!} \sum_{s=0}^{\infty} \frac{\theta^s}{s!}\\ &= \sum_{s=0}^{\infty} \frac{\theta^s}{s!} + \frac{a_n}{1!} \sum_{s=0}^{\infty} \frac{\theta^s}{s!} + \cdots + \frac{a_n^{n-1}}{(n-1)!} \sum_{s=0}^{\infty} \frac{\theta^s}{s!}.\end{aligned}$$

Developing the left-hand side and canceling out the common terms on both sides, we obtain

$$\frac{a_n^n}{n!} = \sum_{s=n+1}^{\infty} \frac{\theta^s}{s!} + \frac{a_n}{1!} \sum_{s=n}^{\infty} \frac{\theta^s}{s!} + \cdots + \frac{a_n^{n-1}}{(n-1)!} \sum_{s=2}^{\infty} \frac{\theta^s}{s!}.$$

Arguing analogously, we derive

$$\frac{b_n^n}{n!} = \sum_{s=n+1}^{\infty} \frac{\delta^s}{s!} + \frac{b_n}{1!} \sum_{s=n}^{\infty} \frac{\delta^s}{s!} + \cdots + \frac{b_n^{n-1}}{(n-1)!} \sum_{s=2}^{\infty} \frac{\delta^s}{s!}.$$

Suppose it were the case that $\delta \geq c$. $b_n/a_n=c$ together with $\delta \geq c\theta$ implies that

$$\begin{aligned}\frac{b_n^n}{n!} &\geq \sum_{s=n+1}^{\infty} \frac{(c\theta)^s}{s!} + \frac{ca_n}{1!} \sum_{s=n}^{\infty} \frac{(c\theta)^s}{s!} + \cdots + \frac{(ac_n)^{n-1}}{(n-1)!} \sum_{s=2}^{\infty} \frac{(c\theta)^s}{s!} \\ &\geq c^{n+1} \sum_{s=n+1}^{\infty} \frac{\theta^s}{s!} + \frac{a_n}{1!} \sum_{s=n}^{\infty} \frac{\theta^s}{s!} + \cdots + \frac{a_n^{n-1}}{(n-1)!} \sum_{s=2}^{\infty} \frac{\theta^s}{s!} \\ &= c^{n+1} \frac{a_n^n}{n!}.\end{aligned}$$

It yields immediately that $c \leq 1$. A contradiction. Therefore, $\delta < c\theta$, and it gives $b_{n+1}/a_{n+1} < b_n/a_n$. This proves the lemma.

3. **Main results.** Population with exponential distribution:

Let $X_1, X_2, \cdots, X_n$ be i.i.d. r.v.'s with common p.d.f. $f(x;\theta) = \theta_e^{-\theta x}$, then it follows that $\bar{X}$-Gamma $(n, 1/n\theta)$. As a consequence, $n\theta\bar{X}$-Gamma $(n, 1)$, and we have

$$P_r(\bar{x} > k_n | \theta = \theta_0) = \int_{n\theta_0 k_n}^{\infty} \frac{t^{n-1} e^{-t}}{(n-1)!} dt,$$

$$P_r(\bar{x} < k_n | \theta = \theta_1) = \int_0^{n\theta_1 k_n} \frac{t^{n-1} e^{-t}}{(n-1)!} dt.$$

THEOREM 1. *If there exists a positive integer n with $b_n/a_n = \theta_0/\theta_1$, then this n is the unique one which makes* (1) *hold.*

Proof. Set $k_n = b_n/n\theta_0$. Then

$$Pr(\bar{x} > k_n | \theta = \theta_0) = \int_{b_n}^{\infty} \frac{t^{n-1} e^{-t}}{(n-1)!} dt = \alpha$$

and

$$Pr(\bar{x} < k_n | \theta = \theta_1) = \int_0^{a_n} \frac{t^{n-1} e^{-t}}{(n-1)!} dt = \beta.$$

Suppose the sample size m also makes (1) hold. This gives that $b_m = m\theta_0 k_m$, $a_m = m\theta_1 k_m$. Hence, $b_m/a_m = \theta_0/\theta_1$. $m = n$ follows from the strictly decreasing property of $\{b_n/a_n\}$.

THEOREM 2. *Suppose there is no integer n with $b_n/a_n = \theta_0/\theta_1$. In this case, the following statements are equivalent.*

(i) *n is the smallest positive integer satisfying $b_n/a_n < \theta_0/\theta_1$.*

(ii) *n is the smallest sample size which makes* (2) *hold.*

(iii) *b_n/a_n is the largest value in the sequence b_k/a_k with $b_n/a_n < \theta_0/\theta_1$.*

Proof. Obviously, sample size n satisfies (2) if and only if $b_n/a_n \leq \theta_0/\theta_1$. It follows immediately from lemma that (i), (ii) and (iii) are equivalent.

COROLLARY. *The smallest sample sizes which make* (2), (7) *and* (8) *hold, respectively, are equal, where*

$$Pr(\bar{x} > k_n | \theta = \theta_0) = \alpha, \quad Pr(\bar{x} < k_n | \theta = \theta_1) < \beta, \tag{7}$$

$$Pr(\bar{x} > k_n | \theta = \theta_0) < \alpha, \quad Pr(\bar{x} < k_n | \theta = \theta_1) = \beta. \tag{8}$$

Proof. Since the smallest sample size n satisfying any one of (2), (7) and (8) must be the smallest positive integer so that $b_n/a_n < \theta_0/\theta_1$.

Tne theorem below gives a lower bound for the smallest sample size n. Its proof need the help of a well-known result, namely, Stirling's formula.

Stirling's formula. $n! = \sqrt{2\pi} e^{-n} n^{n+(1/2)} e^{r(n)/12n}$, where

$$1 - \frac{1}{(12n+1)} < r(n) < 1.$$

THEOREM 3. *If the sample size n makes either* (1) *or* (2) *hold, then* $n > 2\pi((1-\alpha-\beta)/(\ln\theta_0 - \ln\theta_1))^2$.

Proof. Since n makes either (1) or (2) true, it implies that $b_n/a_n \leq \theta_0/\theta_1$. This gives

$$\int_{(\theta_1/\theta_0)b_n}^{b_n} \frac{t^{n-1} e^{-t}}{(n-1)!} \geq \int_{a_n}^{b_n} \frac{t^{n-1} e^{-t}}{(n-1)!} dt$$
$$= 1 - \alpha - \beta.$$

Letting $u = \ln t$, and applying the Mean Value Theorem for Integrals, we find

$$(\ln\theta_0 - \ln\theta_1) \frac{e^{n\tilde{u} - e^{\tilde{u}}}}{(n-1)!} \geq 1 - \alpha - \beta,$$

where $\ln((\theta_1/\theta_0)b_n) < \tilde{u} < b_n$. It can be shown that $u = \ln n$ maximizes $nu - e^u$. Thus we have

$$\frac{n^{n+1}e^{-n}}{n!} \geq \frac{1-\alpha-\beta}{\ln\theta_0 - \ln\theta_1}.$$

Using Stirling's formula, we finally get

$$n > 2\pi\left(\frac{1-\alpha-\beta}{\ln\theta_0 - \ln\theta_1}\right)^2,$$

and the proof is complete.

Population with Poisson distribution:

From now on, assume $X_1, X_2, \cdots, X_n$ are i. i. d. r. v.'s with common p. m. f. $f(x;\theta) = e^{-\theta}\theta^x/x!$, $x = 0, 1, 2, \cdots$. It can be shown that

$$(9) \quad Pr\left(\bar{x} < \frac{k_n}{n}\middle|\theta = \theta_0\right) = \sum_{t=0}^{k_n-1}\frac{e^{-n\theta_0}(n\theta_0)^t}{t!} = \int_{n\theta_0}^{\infty}\frac{t^{k_n-1}e^{-t}}{(k_n-1)!}dt$$

and

$$(10) \quad Pr\left(\bar{x} \geq \frac{k_n}{n}\middle|\theta = \theta_1\right) = 1 - \sum_{t=0}^{k_n-1}\frac{e^{-n\theta_1}(n\theta_1)^t}{t!} = \int_0^{n\theta_1}\frac{t^{k_n-1}e^{-t}}{(k_n-1)!}dt.$$

THEOREM 4. *If there exist positive integers* k_n *and* n *with* $b_{k_n}/\theta_0 = a_{k_n}/\theta_1 = n$, *then* n *is the unique sample size satisfying* (3).

Obviously, the condition in Theorem 4 rarely holds. Now we summarize the key steps for solving the problem in the general case as follows.

(i) Find the smallest positive integer k_0 with $b_{k_0}/a_{k_0} \leq \theta_0/\theta_1$.

(ii) Find the smallest positive integer $k \geq k_0$ so that there exists at least a positive integer between a_k/θ_1 and b_k/θ_0.

(iii) The smallest positive integer between a_k/θ_1 and b_k/θ_0 is the smallest sample size satisfying (4).

A question comes to us naturally. Can (ii) be made? The following theorem gives it an affirmative answer.

THEOREM 5. *We can find a positive integer $k \geq k_0$ such that there exists at least a positive integer between a_k/θ_1 and b_k/θ_0.*

Proof. Suppose, on the contrary, that no integers lie between a_k/θ_1 and b_k/θ_0 for all $k \geq k_0$. Then, for each $k > k_0$, there corresponds an integer n_k with

$$n_k < \frac{b_k}{\theta_0} < \frac{a_k}{\theta_1} < n_k + 1 .$$

The strictly decreasing property of the sequence $\{b_k/a_k\}$ gives

$$1 \leq \frac{\theta_0}{\theta_1} \frac{a_{k_0}}{b_{k_0}} < \frac{\theta_0}{\theta_1} \frac{a_k}{b_k} < 1 + \frac{1}{n_k}$$

for all $k > k_0$. Since $\{a_k\}$ and $\{b_k\}$ are strictly increasing with $\lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k = \infty$, it follows that $\{n_k\}$ is increasing and $\lim_{k\to\infty} n_k = \infty$. Thus, $a_k/b_k = a_{k_0}/b_{k_0}$ for all $k > k_0$. A contradiction. The desired result now follows.

The following theorem justifies the step (iii).

THEOREM 6. *The smallest positive integer choosen in step* (iii), *say n, is the smallest sample size satisfying* (4).

Proof. Suppose the sample size m satisfies (4). It follows immediately from (9) and (10) that $m\theta_0 \geq b_{k_m}$, $m\theta_1 \leq a_{k_m}$. $(b_{k_m}/a_{k_m}) \leq \theta_0/\theta_1$ implies $k_m \geq k_0$. $b_{k_m}/\theta_0 \leq m \leq a_{k_m}/\theta_1$ together with $k_m \geq k_0$ yields $k_m \geq k_n$. Since $b_{k_n}/\theta_0 \leq b_{k_m}/\theta_0$ and n is the smallest integer greater than or equal to b_{k_n}/θ_0, we then must have $m \geq n$.

Similarly, the final theorem below gives a lower bound for the sample size n, whenever n makes either (3) or (4) hold.

THEOREM 7. *If sample size n satisfies either* (3) *or* (4), *then*

$$n \geq \frac{\sqrt{2\pi}(1-\alpha-\beta)}{\theta_0-\theta_1} \sqrt{2\pi\left(\frac{1-\alpha-\beta}{\ln\theta_0-\ln\theta_1}\right)^2 - 1} .$$

Proof. Since n satisfies either (3) or (4), it yields that

$$\int_{n\theta_1}^{n\theta_0} \frac{t^{k_n-1} e^{-t}}{(k_n-1)!} dt \geq 1-\alpha-\beta .$$

It can be shown that $k_n - 1$ maximizes $t^{k_n - 1} e^{-t}$. Hence, we have

$$n \geq \frac{1-\alpha-\beta}{\theta_0-\theta_1} \frac{(k_n-1)!}{(k_n-1)^{k_n-1} e^{-(k_n-1)}}.$$

Using the Stirling's formula, we find

$$n \geq \frac{(1-\alpha-\beta)\sqrt{2\pi(k_n-1)}}{\theta_0-\theta_1}.$$

Since $k_n \geq k_0$ and k_0 is the smallest positive integer with $b_{k_0}/a_{k_0} \leq \theta_0/\theta_1$, it follows from Theorem 3 that

$$k_n > 2\pi\left(\frac{1-\alpha-\beta}{\ln\theta_0-\ln\theta_1}\right)^2.$$

Thus, we conclude that

$$n \geq \frac{\sqrt{2\pi}(1-\alpha-\beta)}{\theta_0-\theta_1}\sqrt{2\pi\left(\frac{1-\alpha-\beta}{\ln\theta_0-\ln\theta_1}\right)^2-1}.$$

Now we illustrate how to use the chi-square table to find out the smallest sample sizes for respective populations. For convenience let $f(x;\ \gamma, \lambda)$ and $F(x;\ \gamma, \lambda)$ denote the p. d. f. and c. d. f. of the distribution Gamma (γ, λ). Here γ, λ are parameters. By changing variables, it is seen easily that $F(x;\ n, 1) = F(2x;\ n, 2)$. Note that $f(x;\ n, 2)$ and $F(x;\ n, 2)$ are, respectively, the p. d. f. and c. d. f. of the chi-square distribution with degrees of freedom $2n$.

Find the smallest degrees of freedom, which is even, from the chi-square table such that $d_l/c_l \leq \theta_0/\theta_1$, where c_l and d_l satisfy $F(c_l;\ n, 2) = \beta$, $F(d_l;\ n, 2) = 1-\alpha$. Then, $n = l/2$ is the smallest sample size for the population with an exponential distribution. If, furthermore, m is the smallest even integer with $d_l/\theta_0 \leq m \leq c_l/\theta_1$, then $n = m/2$ is the smallest sample size for the population Poisson distributed.

Acknowledgements. The author wishes to thank Professor E. S. Thomas and referee for their valuable comments.

References

1. C. R. Rao, *Linear Statistical Inference and its Applications*, John Wiley, New York (1965).

2. V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, Wiley, New York (1976).

Department of Mathematical Sciences
National Chengchi University
Taipei, Taiwan 11623
R. O. C.