⊠연구논문

일반화된 샘플링 계획에서의 가설 검정

김명수 · 오근태
수원대학교 산업공학과

A Hypothesis Test under the Generalized Sampling Plan

Myung-Soo Kim · Geun-Tae Oh
Dept. of Industrial Engineering, Suwon University

Abstract

This paper considers the problem of testing a one-sided hypothesis under the generalized sampling plan which is defined by a sequence of independent Bernoulli trials. A certain lexicographic order is defined for the boundary points of the sampling plan. It is shown that the family of probability mass function defined on the boundary points has monotone likelihood ratio, and that the test function is uniformly most powerful.

1. Introduction

A sequence of independent Bernoulli trials are encountered in many situations such as acceptance sampling by attributes and comparison of two Poisson processes, etc. The parameter of Bernoulli trials is usually unknown and the inference is made from the sample.

The problem of estimating the parameter of Bernoulli trials has been studied by many authors: Haldane(1945), Girshick et al.(1946), Bai et al.(1977), Bai and Kim (1979), Kim and Nachlas(1984), and Kremers(1987) obtained the estimator and studied its properties under various sampling plans. In particular, Kim and Nachlas

(1984) and Kremers(1987) considered the problem under the generalized sampling plan (GSP) which includes fixed size, inverse, fixed size with curtailment, and extended ones. Nachlas and Kim(1989) presented several acceptance sampling plans based on GSP, and Bai and Kim(1993) summarized the theory of statistical inference for the parameter of independent and dependent Bernoulli trials.

Hypothesis testing is a procedure maximizing the power by partitioning the sample space of test statistic into two mutually exclusive and exhaustive subsets, subject to type I error restriction. Since the sample space is determined by a sampling plan, the property of test such as uniformly most powerful (UMP) or uniformly most powerful unbiased (UMPU) etc., depends on the sampling plan. That is, the property of test should be examined under the given sampling plan.

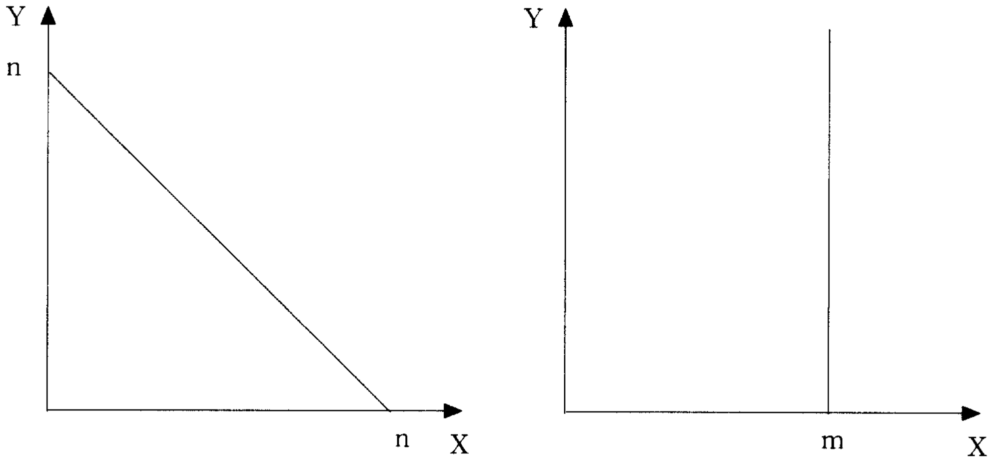
This paper is concerned with the problem of testing a one-sided hypothesis under the generalized sampling plan defined by a sequence of independent Bernoulli trials. By defining a certain lexicographic order for the boundary points of the sampling plan, it is shown that the family of probability mass function defined on the boundary points has monotone likelihood ratio, and that the test function is uniformly most powerful.

2. The Generalized Sampling Plan

A Sequence of Independent Bernoulli Trials

Girshick et al.(1946) presented that a sequence of independent Bernoulli trials can be depicted as a path of sequence of points with nonnegative integral coordinates on the Eucledian plane. That is, paths may be thought of as arising by a random process such that a path reaching (x, y) will be extended to (x+1, y) with probability p or to (x, y+1) with probability 1-p. They divided all points into three mutually exclusive categories; accessible, inaccessible, and boundary points. Accessible points can be reached by paths from the origin, while inaccessible points cannot be reached by any path from the origin. Boundary points are the last points of paths from the origin. The set of boundary points defines a sampling plan which is a rule for terminating sampling, and a sampling plan also determines the set of boundary points. <Figure 1> shows the boundaries of fixed size and inverse sampling plans. Following definition is similar to that of Girshick et al.(1946).

Definition 1. Let B denote the set of boundary points. A sampling plan is said to be closed if and only if $\sum_{x \in B} P(x) = 1$, where $P(x)$ is a probability mass function.



(a) fixed size sampling plan (b) inverse sampling plan

< Figure 1 > The boundaries of fixed size and inverse sampling plans

The Generalized Sampling Plans

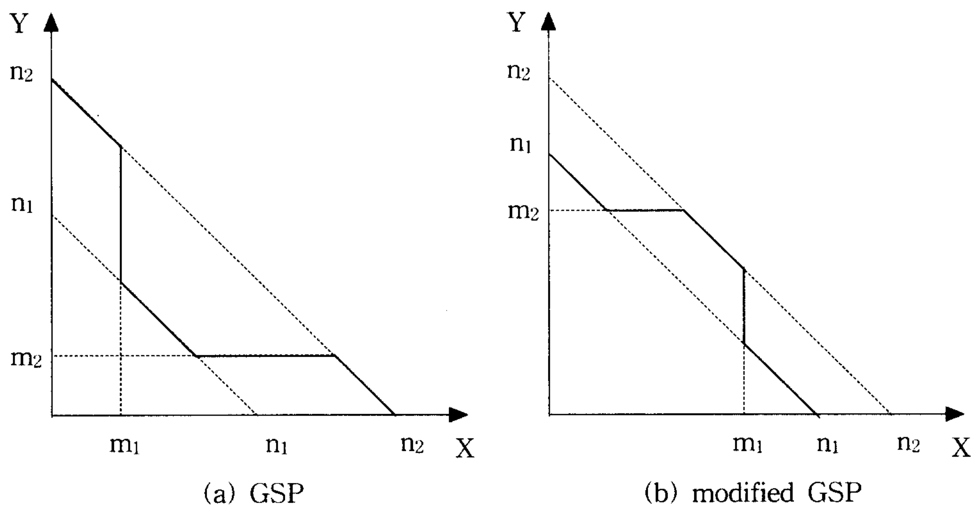
The generalized sampling plan (GSP) proposed by Kim and Nachlas (1984) is a general representation of sampling plans which include fixed size, inverse, fixed size with curtailment, and extended ones. The GSP $S_0(n_1, n_2, m_1, m_2)$ is defined by four parameters, n_1 (the minimum sample size), n_2 (the maximum sample size), m_1 (the number of successes), and m_2 (the number of failures), and its boundary points are as follows:

$$\begin{cases} x+y=n_1, & \text{if } \{\, x \geq m_1 \text{ and } y \geq m_2 \,\} \\ x+y=n_2, & \text{if } \{\, x \leq m_1 \text{ and } y \geq m_2 \,\} \text{ or } \{\, x \geq m_1 \text{ and } y \leq m_2 \,\} \\ x=m_1, & \text{if } \{\, y > m_2 \text{ and } n_1 < x+y < n_2 \,\} \\ y=m_2, & \text{if } \{\, x > m_1 \text{ and } n_1 < x+y < n_2 \,\} \end{cases} \tag{1}$$

Kremers (1987) presented a modified GSP whose stopping rule is:

$$\begin{cases} x+y=n_1, & \text{if } \{ x \geq m_1 \} \text{ or } \{ y \geq m_2 \} \\ x+y=n_2, & \text{if } \{ x \leq m_1,\ y \leq m_2 \text{ and } x+y > n_1 \} \\ x=m_1, & \text{if } \{ y < m_2 \text{ and } n_1 < x+y < n_2 \} \\ y=m_2, & \text{if } \{ x < m_1 \text{ and } n_1 < x+y < n_2 \} \end{cases} \qquad (2)$$

Note that the GSP and the modified GSP are closed sampling plans. <Figure 2> shows the boundaries of the GSP and the modified GSP.



< Figure 2 > The boundaries of a GSP and a modified GSP

3. UMP Test under Generalized Sampling Plan

Let U_i, i=1, 2, ⋯, n, be a sequence of independent Bernoulli trials with parameter p, where $0 \leq p \leq 1$. We now consider the problem of testing the hypothesis of the form under the (modified) GSP:

$$H_0:\ p \leq p_0 \quad \textit{against} \quad H_1:\ p > p_0 \qquad (3)$$

where p_0 is a known constant.

In a (modified) GSP, the boundary point $(X, Y) \in B$ is the sufficient statistic to test the hypothesis (3), where $X = \sum_{i=1}^{n} U_i$ and $Y = n - X$. We now define the lexicographic order for the boundary points (X, Y) of the sampling plan, and monotone likelihood ratio (MLR) for the family of probability mass function defined on the boundary points.

The Lexicographic Order of Boundary Points

Let us define the lexicographic order for the boundary points of a GSP or a modified GSP as:

$$\text{for any} \quad (x, y) \in B, \quad (x_i, y_i) < (x_j, y_j)$$

$$\text{if and only if} \quad (x_i < x_j) \text{ or } (x_i = x_j \text{ and } y_i > y_j).$$

Note that the boundary points of GSP or modified GSP are ordered according to the sequence of boundary line. The boundary points of GSP in Figure 2 is ordered as $(0, n_2) < (m_1, n_2 - m_1) < (m_1, n_1 - m_1) < (n_1 - m_2, m_2) < (n_2 - m_2, m_2) < (n_2, 0)$.

Monotone Likelihood Ratio

Definition 2. Let $\{ f_p,\ p \in [0, 1] \}$ be a family of probability mass function defined on the boundary points of the closed sampling plan. $\{ f_p \}$ is said to have a monotone likelihood ratio in the statistic (X, Y) if $r(x, y) = f_{p_1}(x, y) / f_{p_0}(x, y)$ is a nondecreasing function of $(X, Y) \in B$ for $p_0 < p_1$, whenever f_{p_0} and f_{p_1} are distinct.

Note that $f_p(x, y) = k(x, y) \cdot p^x (1-p)^y$, where $k(x, y)$ is the number of the paths from the origin to (x, y), and $r(x, y) = (p_1 / p_0)^x \cdot [(1-p_1)/(1-p_0)]^y$. The following theorem shows that the family of probability mass function defined on the lexicographically ordered boundary points has monotone likelihood ratio.

Theorem 1. The family of probability mass function defined on the boundary points of GSP which are lexicographically ordered, has monotone likelihood ratio (MRL).

Proof. Suppose that (x_0, y_0) and (x_1, y_1) are boundary points of a (modified) GSP, and assume that $(x_1, y_1) > (x_0, y_0)$. To show that the family of probability mass function defined on boundary points lexicographically ordered has MLR, it is sufficient that the likelihood ratio $r(x, y) = f_{p_1}(x, y) / f_{p_0}(x, y)$ increases in (x, y). Suppose that (x_1, y_1) is adjacent to (x_0, y_0). Then (x_1, y_1) is one of (x_0+1, y_0), (x_0, y_0-1), and (x_0+1, y_0-1). By simple calulations, it can be shown that:

(x_1, y_1)	$r(x_1, y_1)/r(x_0, y_0)$
(x_0+1, y_0)	$p_1/p_0 > 1$
(x_0, y_0-1)	$(1-p_0)/(1-p_1) > 1$
(x_0+1, y_0-1)	$[p_1(1-p_0)]/[p_0(1-p_1)] > 1$

Even if (x_1, y_1) is not adjacent to (x_0, y_0), it can be shown that $r(x_1, y_1)/r(x_0, y_0) > 1$ by sequential applications of the above procedure to adjacent boundary points. Thus $r(x, y)/r(x_0, y_0)$ is increasing in (x, y), and the family of probability mass function defined on the lexicographically ordered boundary points has MLR. □

UMP Test under the Generalized Sampling Plan

The following theorem shows the UMP test for the hypothesis (3) under the generalized sampling plan.

Theorem 2. For testing the hypothesis (3) under the generalized sampling plan, the test function of the form

$$T(x, y) = \begin{cases} 1 & \text{if} \quad (x, y) > (k_1, k_2) \\ \gamma & \text{if} \quad (x, y) = (k_1, k_2) \\ 0 & \text{if} \quad (x, y) < (k_1, k_2) \end{cases} \tag{4}$$

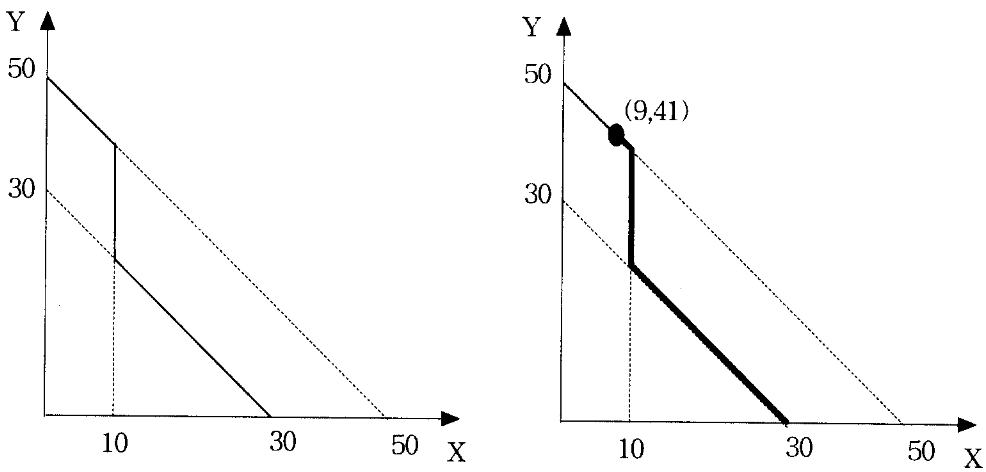
has a nondecreasing power function and is UMP of size $E_{p_0}[T(X, Y)] = \alpha$, where (k_1, k_2) is the lexicographically ordered boundary point.

The proof of Theorm 2 is omitted, since it is similar to that of Lehmann (1986, pp. 78-79).

Example 1. Consider the problem of testing the null hypothesis H_0 : p $\leq$ 0.1 vs. altenative hypothesis H_1 : p > 0.1 under the sampling plan $S_0(30, 50, 10, 0)$. From the Theorem 2, the UMP test of size 0.05 is:

$$T(x, y) = \begin{cases} 1 & \text{if} \quad (x, y) > (9, 41) \\ 0.764 & \text{if} \quad (x, y) = (9, 41) \\ 0 & \textit{otherwise} \end{cases}$$

The sampling plan $S_0(30, 50, 10, 0)$ and the critical region are shown in <Figure 3>.



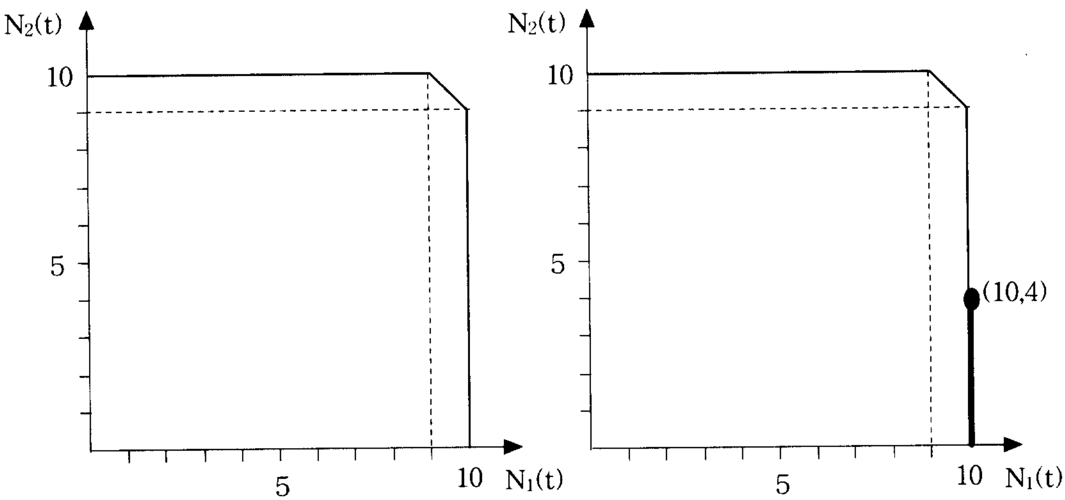
(a) The boundary of $S_0(30, 50, 10, 0)$ (b) The critical region of the test

< Figure 3 > The boundary of S0(30, 50, 10, 0) and critical region for the test

Example 2. Let $\{ N_1(t),\ t \geq 0\}$ and $\{ N_2(t),\ t \geq 0\}$ be the number of events from two independent Poisson processes with occurrence rates λ_1 and λ_2, respectively. Then $(N_1(t),\ N_2(t))$ can be depicted as a path of sequence of points on the Eucledian plane like a sequence of independent Bernoulli trials. That is, a path reaching (x, y) will be extended to (x+1, y) with probability $p = \lambda_1/(\lambda_1 + \lambda_2)$ or to (x, y+1) with probability $1-p=\lambda_2/(\lambda_1+\lambda_2)$. Suppose that the sampling plan $S_0(0, \infty, 10, 10)$ is considered to test the hypothesis H_0: $\lambda_1 \le \lambda_2$ against H_1: $\lambda_1 > \lambda_2$ or equivalently H_0: $\lambda_1/(\lambda_1+\lambda_2) \le 1/2$ against H_1: $\lambda_1/(\lambda_1+\lambda_2) > 1/2$. Note that the sampling plan $S_0(0, \infty, 10, 10)$ terminates observation as soon as 10 events occur from either of two Poisson process. From the Theorem 2, the UMP test of size 0.05 is:

$$T(x,y)=\begin{cases}1 & \text{if} \quad (x,y)>(10,3)\\ 0.0884 & \text{if} \quad (x,y)=(10,4)\\ 0 & otherwise\end{cases}$$

The sampling plan $S_0(0, \infty, 10, 10)$ and the critical region are shown in <Figure 4>.



(a) The boundary of $S_0(0, \infty, 10, 10)$ (b) The critical region of the test

< Figure 4 > The boundary of $S_0(0, \infty, 10, 10)$ and critical region for the test

4. Concluding Remarks

Under the generalized sampling plan defined by a sequence of independent

Bernoulli trials, the UMP test for a one-sided hypothesis is obtained by defining a lexicographic order for boundary points of sampling plan. The result of this paper can be applied to the closed sampling plans depicted in Eucledian plane such as acceptance sampling plan, acceptance sampling plan with curtailment, comparison of two Poisson processes with type II censored data.

References

[1] 배도선, 김성인(1993), *베르누이 시행의 통계적 분석*, 민음사.

[2] Bai, D.S., Kim, S.I., and Lee, J.K.(1977), "On a Generalized Inverse Binomial Sampling Plan," *The Journal of the Korean Statistical Society,* Vol. 4, pp. 1-20.

[3] Bai, D.S. and Kim, S.I.(1979), "Estimation of Transition Probabilities in a Two-State Markov Chain," *Communications in Statistics A,* Vol. 8, pp. 591- 600.

[4] Girshick, M.A., Mosteller, F. and Savage, L.J.(1946), "Unbiased Estimation for Certain Binomial Sampling Problems with Applications," *Annals of Mathematical Statistics,* Vol. 17, pp. 13-23.

[5] Haldane, J.B.C.(1945), "On a Method of Estimating Frequencies," *Biometrika,* Vol. 33, pp. 13-23.

[6] Kim, S.I. and Nachlas, J.A.(1984), "Estimation in Bernoulli Trials under a Generalized Sampling Plan," *Technometrics,* Vol. 26, pp. 379-387.

[7] Kremers, W.K.(1987), "An Improved Estimator of the Mean for Sequential Binomial Sampling Plan," *Technometrics,* Vol. 29, pp. 109-112.

[8] Lehmann, E. L.(1986), *Testing Statistical Hypothesis,* Wiley, New York, 2nd ed.

[9] Nachlas, J.A. and Kim, S.I.(1989), "Generalized Attribute Acceptance Sampling Plans," *Journal of Quality Technology*, Vol. 21, pp. 32-40.