한국통계학회논문집 제2권 1호
1995년 4월 pp. 112-121

Comparison Density Representation of Traditional Test Statistics for the Equality of Two Population Proportions

Jangsun Baek[1]

Abstract

Let p_1 and p_2 be the proportions of two populations. To test the hypothesis $H_0: p_1 = p_2$, we usually use the χ^2 statistic, the large sample binomial statistic Z, and the Generalized Likelihood Ratio statistic $-2\log\lambda$, which were developed based on different mathematical rationale, respectively. Since testing the above hypothesis is equivalent to testing whether two populations follow the common Bernoulli distribution, one may also test the hypothesis by comparing 1 with the ratio of each density estimate and the hypothesized common density estimate, called comparison density, which was devised by Parzen(1988). We show that the above traditional test statistics are actually estimating the measure of distance between the true densities and the common density under H_0 by representing them with the comparison density.

1. Introduction

We often present binomial data gathered from more than one population in a contingency table. For the case of two populations, suppose X_1 is the number of successes in population 1, $X_1 \sim$ BIN(n_1, p_1) with the realization of n_{11} successes and n_{12} failures, and X_2 is the number of successes in population 2, $X_2 \sim$ BIN(n_2, p_2) with the realizations of n_{21} successes and n_{22} failures. Then the contingency table might look like as in Table 1.

Table 1. 2×2 Contingency table

	Successes	Failures	Total
Population 1 (X_1)	n_{11}	n_{12}	n_1
Population 2 (X_2)	n_{21}	n_{22}	n_2
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	n

1) Department of Statistics, Chonnam National University, Kwangju, 500-757, KOREA

A hypothesis that is usually of interest is

$$H_0 : p_1 = p_2 \quad \text{versus} \quad H_1 : p_1 \neq p_2 \quad . \tag{1}$$

There are several test procedures we can use to test the above hypotheses. The χ^2 test, the large sample binomial test, and the Generalized Likelihood Ratio (GLR) test, which are described in Section 2, might be the most widely used traditional procedures (Bain and Engelhardt, 1992).

Testing the hypotheses of (1) is equivalent to testing the equality of two Bernoulli probability density functions (p.d.f.). More specifically, for a Bernoulli random variable Y, let $f_i(y) = p_i I_{\{z_1\}}(y) + (1-p_i) I_{\{z_2\}}(y)$ be the p.d.f. of Y for the ith population, where $I_{\{z_i\}}(y)$ is the indication function, $i=1,2$. Then X_i can be regarded as the sum of the random sample $\{ Y_{i1}, Y_{i2}, \cdots, Y_{in_i} \}$ from the ith population, $i=1,2$. When $p_1 = p_2 = p$, the pooled sample $\{ Y_{11}, Y_{12}, \cdots, Y_{1n_1}, \ Y_{21}, Y_{22}, \cdots, Y_{2n_2} \}$ of size $n = n_1 + n_2$ is regarded as observations on a variable Y with the common p.d.f., $f(y) = p I_{\{z_1\}}(y) + (1-p) I_{\{z_2\}}(y)$. Thus the hypotheses of (1) can be restated as

$$H_0 : f_1(y) = f_2(y) = f(y) \quad \text{versus} \quad H_1 : f_1(y) \neq f_2(y) \ , \tag{2}$$

where $f(y)$ is the common Bernoulli p.d.f. with the probability of success $p = p_1 = p_2$. Therefore the test for the equality of two population proportions is just a special case of the test for the homogeneity of two distributions.

Based on the entropy, $-\int g(x) \log g(x) dx$ defined by Shannon(1948) as a measure of uncertainty of a random variable X with its p.d.f. $g(x)$, Kullback(1959) defined the relative entropy (cross-entropy or Kullback Leibler distance) as $\int g(x) \log \{ g(x)/h(x) \} dx$ to measure the distance between two p.d.f.'s $g(x)$ and $h(x)$. The relative entropy for discrete distribution is $D(g \| h) = \sum_x g(x) \log \{ g(x)/h(x) \}$. It is easy to show using Jensen's inequality that $D(g \| h) \geq 0$ with equality if and only if $g(x) = h(x)$ for all x. See Cover and Thomas (1992) for more about the relationship between information theory and Statistics.

Direct comparison of the ratio of two densities, $g(\cdot)/h(\cdot)$, in the relative entropy with 1 can be used to construct a test statistic for testing the hypotheses of (2). One may measure the distance between the two true p.d.f.'s f_1, f_2 and the common p.d.f., f under H_0 by a functional of the two ratios, f_1/f and f_2/f. The index i of an observation in the pooled sample $\{ \{ Y_{i1}, Y_{i2}, \cdots, Y_{in_i} \}_{i=1}^{2} \}$ is regarded as the value of a variable W. The sample probability that $W = i$ is denoted $\lambda_i = n_i/n$, $i=1,2$. One forms the sample p.d.f. of Y_i, $\hat{f}_i(z_j) = n_{ij}/n_i$ which estimates the true f_i's under the alternative hypothesis H_1 , $i=1,2$, $j=1,2$. Under the null hypothesis H_0 , the pooled sample probability that $Y = z_j$ denoted by $\hat{f}(z_j) = n_{\cdot j}/n$, estimates the common p.d.f., f, $j=1,2$. Parzen(1988) defined the comparison density, $d(u, (\hat{f}, \hat{f}_i))$ as

$$d_i(u, (\hat{f}, \hat{f}_i)) = \{ \hat{f}_i(z_1)/\hat{f}(z_1) \} I_{\{0<u<\hat{f}(z_1)\}}(u) + \{ \hat{f}_i(z_2)/\hat{f}(z_2) \} I_{\{\hat{f}(z_1)<u<1\}}(u), \; i=1,2, \quad (3)$$

and proposed a test statistic which is a functional of $d_i(u, (\hat{f}, \hat{f}_i))$,

$$C = \sum_{i=1}^{2} \lambda_i \int_0^1 \{ d_i(u, (\hat{f}, \hat{f}_i)) - 1 \}^2 du \quad (4)$$

to test the hypotheses of (2). As the comparison density measures the distance between the two densities, f and f_i, we can define the comparison density differently, such as $d_i(u, (\hat{f}_i, \hat{f})) = \{ \hat{f}(z_1)/\hat{f}_i(z_1) \} I_{\{0<u<\hat{f}_i(z_1)\}}(u) + \{ \hat{f}(z_2)/\hat{f}_i(z_2) \} I_{\{\hat{f}_i(z_1)<u<1\}}(u)$. See the examples of $d_i(u, (\hat{f}, \hat{f}_i))$ and $d_i(u, (\hat{f}_i, \hat{f}))$ in Figure 1. Note that $d_i(u, (\hat{f}, \hat{f}_i))$ is a density function. That is $d_i(u, (\hat{f}, \hat{f}_i)) \geq 0$ for $0 < u < 1$ and $\int d_i(u, \hat{f}, \hat{f}_i))du = 1$. As f_i is similar to f, $d_i(u, (\hat{f}, \hat{f}_i))$ gets close to 1 and C becomes small, but C becomes large, otherwise. Though he suggested various versions of test statistics based on appropriately defined comparison densities in other testing situations, we focus on the comparison density of (3) for two sample discrete data, and want to show that some traditional test procedures can be explained in terms of the comparison density.

In section 2 we summarize the traditional test procedures. In the final section we show the relationship between the traditional test statistics and the comparison density of (3). It is shown that the χ^2 test statistic and the large sample binomial test statistic are basically the same as C in (4), and the GLR test statistic is a functional of the comparison density

$d_i(u;(\hat{f}_i, \hat{f}))$.

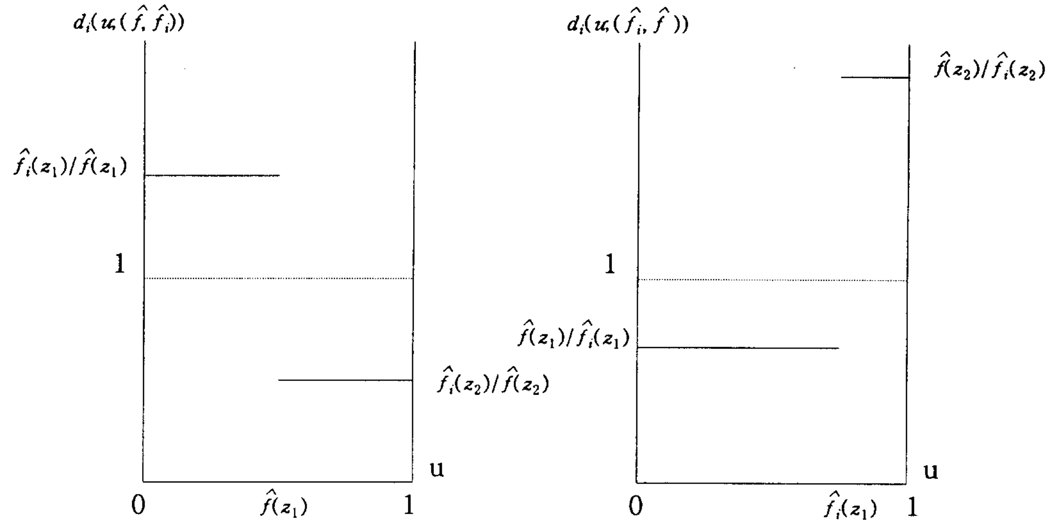

Figure 1. Examples of $d_i(u;(\hat{f}, \hat{f}_i))$ and $d_i(u;(\hat{f}_i, \hat{f}))$

2. Traditional test statistics

The most popular test statistic to test the hypotheses of (1) for 2×2 contingency table data would be the χ^2 statistic. That is,

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(O_{ij}-\hat{E}_{ij})^2}{\hat{E}_{ij}} \quad , \tag{5}$$

where O_{ij} is the observed outcomes of Z_j in the ith sample, and $\hat{E}_{ij}$ is the estimated expected outcomes under H_0. We know $O_{ij}=n_{ij}$ and $\hat{E}_{ij}=n_i n_{\cdot j}/n$ in this test, $i=1,2,\ j=1,2$. The above χ^2 statistic is approximately distributed as $\chi^2_{(1)}$ under H_0 .

It is also possible to construct test of the hypotheses using large sample theory. The maximum likelihood estimators (m.l.e.) of p_1, p_2 are $\hat{p}_1=n_{11}/n_1$ and $\hat{p}_2=n_{21}/n_2$, respectively. Under H_0: $p_1=p_2$, it would seem appropriate to have a pooled estimator of their common value, $\hat{p}=(n_{11}+n_{21})/n=n_{\cdot 1}/n$. Applying large sample theory to these estimators we can construct the following large sample binomial test statistic Z, which is approximately distributed as $N(0,1)$ under H_0 ;

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}} \quad . \tag{6}$$

Let the parameter space be $\Omega = \{\, \theta = (p_1, p_2) | 0 < p_1 < 1 \text{ and } 0 < p_2 < 1 \,\}$, and let the subset corresponding to H_0 be $\Omega_0 = \{\theta = (p_1, p_2) | 0 < p_1 = p_2 < 1\}$. Based on the binomial data, the m.l.e.'s are $\hat{p}_1 = n_{11}/n_1$, $\hat{p}_2 = n_{21}/n_2$ over Ω and $\hat{p} = n_{\cdot 1}/n$ over Ω_0. Then the GLR statistic is

$$\lambda = \frac{\max_{\theta \in \Omega_0} L(x_1, x_2; \theta)}{\max_{\theta \in \Omega} L(x_1, x_2; \theta)}$$

$$= \frac{\binom{n_1}{n_{11}} \hat{p}^{n_{11}} (1-\hat{p})^{n_{12}} \binom{n_2}{n_{21}} \hat{p}^{n_{21}} (1-\hat{p})^{n_{22}}}{\binom{n_1}{n_{11}} \hat{p}_1^{n_{11}} (1-\hat{p}_1)^{n_{12}} \binom{n_2}{n_{21}} \hat{p}_2^{n_{21}} (1-\hat{p}_2)^{n_{22}}} \tag{7}$$

Since $-2\log\lambda$ is approximately distributed as $\chi^2_{(1)}$ under H_0 for large sample, the commonly used GLR statistic is $-2\log\lambda$.

3. Main results

Now we study the relationship between the conventional test statistics and the comparison density. It is shown in the following proposition that the traditional test statistics for binomial data, χ^2, Z, and $-2\log\lambda$ can be expressed in terms of the comparison density, and thus actually compare the ratio of the estimates of two densities under the null and alternative hypotheses to 1.

Proposition Let p_1 and p_2 be the proportions of two different populations. For the binomial data given in Table 1, suppose χ^2, Z are the χ^2 test statistic, the large sample binomial test statistic defined in (5), (6), respectively, and $-2\log\lambda$ is the GLR test statistic based on λ defined in (7) to test the hypotheses: $H_0 : p_1 = p_2$ versus $H_1 : p_1 \neq p_2$. Let $d_i(u;(\hat{f},\hat{f}_i)) = \{\hat{f}_i(z_1)/\hat{f}(z_1)\}I_{\{0<u<\hat{F}(z_1)\}}(u) + \{\hat{f}_i(z_2)/\hat{f}(z_2)\}I_{\hat{F}(z_1)<u<1\}}(u)$, and let $d_i(u;(\hat{f}_i,\hat{f})) = \{\hat{f}(z_1)/\hat{f}_i(z_1)\}I_{\{0<u<\hat{F}_i(z_1)\}}(u) + \{\hat{f}(z_2)/\hat{f}_i(z_2)\}I_{\{\hat{F}_i(z_1)<u<1\}}(u)$. Then

$$\text{(i)} \quad \chi^2 = n\sum_{i=1}^{2}\lambda_i\int_0^1\{d_i(u;\hat{f},\hat{f}_i)) - 1\}^2 du$$

$$\text{(ii)} \quad Z^2 = n\sum_{i=1}^{2}\lambda_i\int_0^1\{d_i(u;(\hat{f},\hat{f}_i)) - 1\}^2 du$$

$$\text{(iii)} \quad -2\log\lambda = 2n\sum_{i=1}^{2}\lambda_i\int_0^1 -\log d_i(u;(\hat{f}_i,\hat{f}))du ,$$

where $\lambda_i = n_i/n$, i=1,2.

Proof (i) Note $O_{ij} = n_{ij}$ and $\hat{E}_{ij} = n_i n_{\cdot j}/n$, then

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij} - \frac{n_i n_{\cdot j}}{n}\right)^2}{\frac{n_i n_{\cdot j}}{n}} = \sum_{i=1}^{2}\sum_{j=1}^{2}\left\{\frac{n_i n_{\cdot j}}{n}\Big/\left(\frac{n_i n_{\cdot j}}{n}\right)^2\right\}\left(n_{ij} - \frac{n_i n_{\cdot j}}{n}\right)^2$$

$$= \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{n_i n_{\cdot j}}{n}\left(n_{ij}\Big/\left(\frac{n_i n_{\cdot j}}{n}\right) - 1\right)^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{n_i n_{\cdot j}}{n}\left(\frac{n_{ij}/n_i}{n_{\cdot j}/n} - 1\right)^2$$

$$= n\sum_{i=1}^{2}\frac{n_i}{n}\sum_{j=1}^{2}\frac{n_{\cdot j}}{n}\left(\frac{n_{ij}/n_i}{n_{\cdot j}/n} - 1\right)^2 = n\sum_{i=1}^{2}\lambda_i\sum_{j=1}^{2}\hat{f}(z_j)\left(\frac{\hat{f}_i(z_j)}{\hat{f}(z_j)} - 1\right)^2 ,$$

since $\lambda_i = n_i/n$, $\hat{f}_i(z_j) = n_{ij}/n_i$, and $\hat{f}(z_j) = n_{\cdot j}/n$. Thus the last equation is equal to $n\sum_{i=1}^{2}\lambda_i\int_0^1\{\, d_i(u,(\hat{f},\hat{f}_i) - 1\,\}^2 du$.

(ii) We square the statistic Z. Then also by noting $1/\hat{p} + 1/\hat{q} = 1/(\hat{p}\hat{q})$,

$$Z^2 = \frac{(\hat{p}_1 - \hat{p}_2)^2}{(1/n_1 + 1/n_2)\hat{p}\hat{q}} = \left(\frac{n_1 n_2}{n_1 + n_2}\right)\frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}} + \left(\frac{n_1 n_2}{n_1 + n_2}\right)\frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{q}} .$$

Note $n_1 + n_2 = n$, and $(n_1 n_2)/(n_1+n_2) = n(n_1/n)(n_2/n) = n\lambda_1\lambda_2$. Thus $Z^2 = n\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p}_2)^2}{\hat{p}} + n\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p}_2)^2}{\hat{q}}$. Divide both sides by n .

$$\frac{Z^2}{n} = \lambda_1\lambda_2\frac{\{(\hat{p}_1-\hat{p})-(\hat{p}_2-\hat{p})\}^2}{\hat{p}} + \lambda_1\lambda_2\frac{\{(\hat{p}_1-\hat{p})-(\hat{p}_2-\hat{p})\}^2}{\hat{q}}$$

$$= \lambda_1\lambda_2\frac{\{(\hat{p}_1-\hat{p})^2+(\hat{p}_2-\hat{p})^2-2(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})\}}{\hat{p}}$$

$$+ \lambda_1\lambda_2\frac{\{(\hat{p}_1-\hat{p})^2+(\hat{p}_2-\hat{p})^2-2(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})\}}{\hat{q}}$$

$$= \lambda_1\lambda_2\left\{\frac{(\hat{p}_1-\hat{p})^2+(\hat{p}_2-\hat{p})^2}{\hat{p}}\right\} + \lambda_1\lambda_2\left\{\frac{(1-\hat{q}_1-1+\hat{q})^2+(1-\hat{q}_2-1+\hat{q})^2}{\hat{q}}\right\}$$

$$- 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}$$

$$= \lambda_1\lambda_2\left\{\frac{(\hat{p}_1-\hat{p})^2+(\hat{p}_2-\hat{p})^2}{\hat{p}}\right\} + \lambda_1\lambda_2\left\{\frac{(\hat{q}_1-\hat{q})^2+(\hat{q}_2-\hat{q})^2}{\hat{q}}\right\}$$

$$- 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}} .$$

Since $\left(\frac{\hat{p}_i}{\hat{p}}-1\right)^2 = \left(\frac{\hat{p}_i-\hat{p}}{\hat{p}}\right)^2 = \frac{1}{\hat{p}^2}(\hat{p}_i-\hat{p})^2$, $(\hat{p}_i-\hat{p})^2 = \hat{p}^2\left(\frac{\hat{p}_i}{\hat{p}}-1\right)^2$, $i=1,2$.

Similarly $(\hat{q}_i-\hat{q})^2 = \hat{q}^2\left(\frac{\hat{q}_i}{\hat{q}}-1\right)^2$, $i=1,2$. Then the last equation becomes

$$\lambda_1\lambda_2\left\{\hat{p}\left(\frac{\hat{p}_1}{\hat{p}}-1\right)^2+\hat{p}\left(\frac{\hat{p}_2}{\hat{p}}-1\right)^2+\hat{q}\left(\frac{\hat{q}_1}{\hat{q}}-1\right)^2+\hat{q}\left(\frac{\hat{q}_2}{\hat{q}}-1\right)^2\right\}$$

$$- 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}$$

$$= \lambda_1(1-\lambda_1)\left\{ \hat{p}\left(\frac{\hat{p}_1}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_1}{\hat{q}}-1\right)^2 \right\} + \lambda_2(1-\lambda_2)\left\{ \hat{p}\left(\frac{\hat{p}_2}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_2}{\hat{q}}-1\right)^2 \right\}$$

$$- 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}$$

$$= \lambda_1\left\{ \hat{p}\left(\frac{\hat{p}_1}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_1}{\hat{q}}-1\right)^2 \right\} + \lambda_2\left\{ \hat{p}\left(\frac{\hat{p}_2}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_2}{\hat{q}}-1\right)^2 \right\}$$

$$-\left[\lambda_1^2\left\{ \hat{p}\left(\frac{\hat{p}_1}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_1}{\hat{q}}-1\right)^2 \right\} + 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}\right.$$

$$\left. + \lambda_2^2\left\{ \hat{p}\left(\frac{\hat{p}_2}{\hat{p}}-1\right)^2 + \hat{q}\left(\frac{\hat{q}_2}{\hat{q}}-1\right)^2 \right\}\right] .$$

Consider the expression in the bracket, and let it be A, then

$$A = \lambda_1^2\left\{ \frac{(\hat{p}_1-\hat{p})^2}{\hat{p}} + \frac{(\hat{q}_1-\hat{q})^2}{\hat{q}} \right\} + 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}$$

$$+ \lambda_2^2\left\{ \frac{(\hat{p}_2-\hat{p})^2}{\hat{p}} + \frac{(\hat{q}_2-\hat{q})^2}{\hat{q}} \right\}$$

$$= \lambda_1^2\left\{ \frac{(\hat{p}_1-\hat{p})^2}{\hat{p}} + \frac{(\hat{p}_1-\hat{p})^2}{\hat{q}} \right\} + 2\lambda_1\lambda_2\frac{(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p})}{\hat{p}\hat{q}}$$

$$+ \lambda_2^2\left\{ \frac{(\hat{p}_2-\hat{p})^2}{\hat{p}} + \frac{(\hat{p}_2-\hat{p})^2}{\hat{q}} \right\}$$

$$= \frac{1}{\hat{p}\hat{q}}\{ \lambda_1^2(\hat{p}_1-\hat{p})^2 + 2\lambda_1\lambda_2(\hat{p}_1-\hat{p})(\hat{p}_2-\hat{p}) + \lambda_2^2(\hat{p}_2-\hat{p})^2 \}$$

$$= \frac{1}{\hat{p}\hat{q}}\{ \lambda_1(\hat{p}_1-\hat{p}) + \lambda_2(\hat{p}_2-\hat{p}) \}^2$$

$$= \frac{1}{\hat{p}\hat{q}}\left\{ \frac{n_1}{n}\left(\frac{n_{11}}{n_1} - \frac{n_{\cdot 1}}{n}\right) + \frac{n_2}{n}\left(\frac{n_{21}}{n_2} - \frac{n_{\cdot 1}}{n}\right)\right\}$$

$$= \frac{1}{\hat{p}\hat{q}}\left\{ \frac{n_{11}}{n} - \frac{n_1 n_{\cdot 1}}{n^2} + \frac{n_{21}}{n} - \frac{n_2 n_{\cdot 1}}{n^2} \right\}$$

$$= \frac{1}{\hat{p}\hat{q}} \left\{ \frac{n_{\cdot 1}}{n} - \frac{n\, n_{\cdot 1}}{n^2} \right\}$$

$$= 0 .$$

Thus $\frac{Z^2}{n} = \lambda_1 \left\{ \hat{p}\left(\frac{\hat{p}_1}{\hat{p}} - 1\right)^2 + \hat{q}\left(\frac{\hat{q}_1}{\hat{q}} - 1\right)^2 \right\} + \lambda_2 \left\{ \hat{p}\left(\frac{\hat{p}_2}{\hat{p}} - 1\right)^2 + \hat{q}\left(\frac{\hat{q}_2}{\hat{q}} - 1\right)^2 \right\}$.

Remember $\hat{p}_1 = n_{11}/n_1 = \hat{f}_1(z_1)$, $\hat{q}_1 = n_{12}/n_1 = \hat{f}_1(z_2)$, $\hat{p}_2 = n_{21}/n_2 = \hat{f}_2(z_1)$, $\hat{q}_2 = n_{22}/n_2 = \hat{f}_2(z_2)$, $\hat{p} = n_{\cdot 1}/n = \hat{f}(z_1)$, and $\hat{q} = n_{\cdot 2}/n = \hat{f}(z_2)$. Hence from the last equation above,

$$Z^2 = n \sum_{i=1}^{2} \lambda_i \int_0^1 \{ d_i(u; (\hat{f}, \hat{f}_i) - 1 \}^2 du .$$

(iii) From the expression (7), we obtain

$$\frac{\log \lambda}{n} = \frac{n_{11}}{n} \log \frac{\hat{p}}{\hat{p}_1} + \frac{n_{12}}{n} \log \frac{\hat{q}}{\hat{q}_1} + \frac{n_{21}}{n} \log \frac{\hat{p}}{\hat{p}_2} + \frac{n_{22}}{n} \log \frac{\hat{q}}{\hat{q}_2}$$

$$= \left(\frac{n_1}{n}\right)\left(\frac{n_{11}}{n_1}\right) \log \frac{\hat{p}}{\hat{p}_1} + \left(\frac{n_1}{n}\right)\left(\frac{n_{12}}{n_1}\right) \log \frac{\hat{q}}{\hat{q}_1} + \left(\frac{n_2}{n}\right)\left(\frac{n_{21}}{n_2}\right) \log \frac{\hat{p}}{\hat{p}_2}$$

$$+ \left(\frac{n_2}{n}\right)\left(\frac{n_{22}}{n_2}\right) \log \frac{\hat{q}}{\hat{q}_2}$$

$$= \frac{n_1}{n} \left\{ \hat{p}_1 \log \frac{\hat{p}}{\hat{p}_1} + \hat{q}_1 \log \frac{\hat{q}}{\hat{q}_1} \right\} + \frac{n_2}{n} \left\{ \hat{p}_2 \log \frac{\hat{p}}{\hat{p}_2} + \hat{q}_2 \log \frac{\hat{q}}{\hat{q}_2} \right\}$$

$$= \lambda_1 \left\{ \hat{f}_1(z_1) \log \frac{\hat{f}(z_1)}{\hat{f}_1(z_1)} + \hat{f}_1(z_2) \log \frac{\hat{f}(z_2)}{\hat{f}_1(z_2)} \right\}$$

$$+ \lambda_2 \left\{ \hat{f}_2(z_1) \log \frac{\hat{f}(z_1)}{\hat{f}_2(z_1)} + \hat{f}_2(z_2) \log \frac{\hat{f}(z_2)}{\hat{f}_2(z_2)} \right\}$$

$$= \sum_{i=1}^{2} \lambda_i \int_0^1 \log d_i(u; (\hat{f}_i, \hat{f})) du .$$

Multiplying −2 on both sides, the result is obtained.

References

[1] Bain, L. J. and Engelhardt, M.(1992). *Introduction to Probability and Mathematical Statistics*, PWS-KENT, Boston.

[2] Cover, T. M. and Thomas, J. A. (1992). *Elements of Information Theory,* John Wiley & Sons.

[3] Kullback, S. (1959). *Information Theory and Statistics,* Wiley, New York.

[4] Parzen, E.(1988). *Muluti-sample Functional Statistical Data Analysis,* Manuscript, Texas A&M University, College Station.

[5] Shannon, C. E. (1948). A Mathematical Theory of Communication, *Bell Systems Technical Journal* ,Vol.27, 379-423 ,623-656.