A study memo on chi-squared test for categorical data

Contents

1 Chi-squared test for categorical data **2**

1 Chi-squared test for categorical data

Proposition 1.1. *Let*

- *(S1)* $(\Omega, \mathscr{F}, P)$ *is a probability space.*
- *(S2)* $\{X_i\}_{i=1}^{\infty}$ *is a sequence of* N*-dimensional vectors of random variables on* $(\Omega, \mathscr{F}, P)$*.*
- *(A1)* $\{X_i\}_{i=1}^{\infty}$ *distribution converges to* $N(0, E_N)$*.*

then $\{|X_i|^2\}_{i=1}^{\infty}$ *distribution converges to* $\chi^2(N)$*.*

Proof. Let us fix arbitary $a > 0$.

Let λ be the N-dimensional Lebesugue's measure. By (A1) and $\lambda(\partial B(X, \sqrt{a})) = 0$,

$$\begin{aligned} \mu(\{|X_i|^2 \leq a\}) &= \mu(\{X_i \in \overline{B(X, \sqrt{a})}\}) \\ &\to N(0, E_N)(\overline{B(X, \sqrt{a})}) \ (i \to \infty) \end{aligned} \tag{1}$$

By the definition of chi-squared distribution with degree of free N,

$$N(0, E_N)(\overline{B(X, a)}) = \chi^2(N)([0, a]) \tag{2}$$

So $\{|X_i|^2\}_{i=1}^{\infty}$ distribution converges to $\chi^2(N)$. □

Theorem 1.1. *Let*

- *(S1)* $(\Omega, \mathscr{F}, P)$ *is a probability space.*
- *(S2)* $\{X_i\}_{i=1}^{\infty}$ *is a sequence of* K*-dimensional vectors of random variables on* $(\Omega, \mathscr{F}, P)$*.*
- *(S3)* $\{\pi_k\}_{k=1}^{K} \subset (0, 1)$ *such that* $\Sigma_{k=1}^{K} \pi_k = 1$*.*
- *(A1)* $P(\{X_{i,k} = 1\}) = 1 \ (\forall i, \forall k)$*.*
- *(A2) For any* k, l *such that* $k \neq l$*,* $\{X_{i,k} = 1\} \cup \{X_{i,l} = 1\} = \phi \ (\forall i)$*.*
- *(S4)* $O_{n,k} := \Sigma_{i=1}^{n} X_{i,k} \ (n \in \mathbb{N}, k \in \mathbb{N})$*.*
- *(S5)* $E_{n,k} := n\pi_k \ (n \in \mathbb{N}, k \in \mathbb{N})$*.*

then

$$Q(n) := \Sigma_{k=1}^{K} \frac{(O_{n,k} - E_{n,k})^2}{n\pi_i} \tag{3}$$

distribution converges to $\chi^2(K-1)$*.*

Proof. We set

$$Y_{n,k} := \sqrt{n}(\bar{X}_k - \pi_k) \ (n \in \mathbb{N}, k \in \mathbb{N}) \tag{4}$$

Then

$$Y_{n,K} := -\Sigma_{k=1}^{K-1} Y_{n,k} \ (\forall n) \tag{5}$$

and

$$O_{n,k} - E_{n,k} = \sqrt{n} Y_{n,k} \ (n \in \mathbb{N}, k \in \mathbb{N}) \tag{6}$$

$$Y_n := (Y_{n,1}, ..., Y_{n,K-1})^T \tag{7}$$

If we set $A := \{a_{i,j}\}_{i,j=1,...,K-1}$ by

$$a_{i,j} = \begin{cases} \frac{1}{\pi_i} + \frac{1}{\pi_K} & \text{if } (i = j), \\ \frac{1}{\pi_K} & \text{if } (i \neq j), \end{cases} \tag{8}$$

So

$$Q(n) = Y_n^T A Y_n \ (n \in \mathbb{N}) \tag{9}$$

and A is a nonnogetive definite symmetric matrix.

We set $(K-1)$-by-$(K-1)$ matrix $\Sigma := \{\sigma_{i,j}\}_{i,j=1,...,K-1}$ by $\sigma_{i,j} = cov(X_{1,i}, X_{1,j})$. Then

$$\sigma_{i,j} = \begin{cases} \pi_i(1-\pi_i) & \text{if } (i = j), \\ -\pi_i \pi_j & \text{if } (i \neq j), \end{cases} \tag{10}$$

and

$$\sigma_{i,j} = cov(X_{n,i}, X_{n,j}) \ (\forall n, \forall i, \forall j) \tag{11}$$

By Proposition1.2, Σ is positive definite symmetric matrix.

By the central limit theomre(see [3]), $Y_{n n=1}^{\infty}$ distribution converges to $N(0, \Sigma)$.

By Proposition1.1, $\{Q(n)\}_{n=1}^{\infty}$distribution converges to $\chi^2(K-1)$.

□

Proposition 1.2. *Let A and B be matrixies in the proof oh Theorem1.1. Then* $A^{-1} = \Sigma$

Proof. For any $i \in \{1, 2, ..., K-1\}$

$$\begin{aligned} (A\Sigma)_{i,i} &= a_{i,i}\sigma_{i,i} + \Sigma_{k\neq i} a_{i,k}\sigma_{k,i} \\ &= (\frac{1}{\pi_i} + \frac{1}{\pi_K})\pi_i(1-\pi_i) + \Sigma_{k\neq i}\frac{1}{\pi_K}(-\pi_i\pi_j) \\ &= (1-\pi_i) + \pi_i\frac{(1-\pi_i) - \Sigma_{k\neq i}\pi_k}{\pi_K} \\ &= 1 \end{aligned} \tag{12}$$

For any $i \in \{1, 2, ..., K-1\}$ and any $j \in \{1, 2, ..., K-1\}$ such that $i \neq j$,

$$\begin{aligned} (A\Sigma)_{i,j} &= a_{i,i}\sigma_{i,j} + a_{i,j}\sigma_{j,j} + \Sigma_{k\neq i,j} a_{i,k}\sigma_{k,i} \\ &= (\frac{1}{\pi_i} + \frac{1}{\pi_K})(-\pi_i\pi_j) + \frac{1}{\pi_K}\pi_j(1-\pi_j) + \Sigma_{k\neq i,j}\frac{1}{\pi_K}(-\pi_k\pi_j) \\ &= (-\pi_j - \frac{\pi_j}{\pi_K}\pi_i) + (\frac{\pi_j}{\pi_K} - \frac{\pi_j}{\pi_K}\pi_j) - \frac{\pi_j}{\pi_K}\Sigma_{k\neq i,j}\pi_k \\ &= -\pi_j + \frac{\pi_j}{\pi_K} - \frac{\pi_j}{\pi_K}(1-\pi_K) \\ &= 0 \end{aligned} \tag{13}$$

□

References

[1] Tadahisa Funaki, Probability Theory(in Japanese), ISBN-13 978-4254116007.

[2] Shinichi Kotani, Measure and Probability(in Japanese), ISBN4-00-010634-1.

[3] A study memo on a proof of the central limit theorem https://osmanthus.work/?p=607

[4] Tatsuya Kubota, Foundations of modern mathematical statistics(in Japanese), ISBN978-4-320-11166-0.