Ch.7 Test

7.1. FORMAL LOGIC OF HYPOTHESIS TESTING

Let us consider the process of exploring a model for a random phenomenon. In such situations, a model is formulated based on experience or prior knowledge, and we examine how well the observed results conform to the proposed model or how much they deviate from it. In particular, if we obtain an observation that would be extremely unlikely under the proposed model, we reject that model and instead seek an alternative that can replace it. According to this logic, the procedure of deciding whether to "accept" the proposed model or to "reject" it and consider alternatives is called statistical hypothesis, or simply testing.

Ex. 7.1.1

A painkiller produced by a pharmaceutical company is known to have a time to onset of analgesic effect that follows a normal distribution with a mean of 30 minutes and a standard deviation of 5 minutes. Researchers at the company developed a new painkiller and claim that its analgesic effect occurs more quickly than that of the existing product. To verify this claim, the company's management plans to randomly select 100 patients, administer the new painkiller to them, and observe the sample mean $\bar{x}$ of the time until the analgesic effect appears. If, based on this study, there is clear evidence that the new painkiller acts more quickly, the company intends to modify its production process to mass-produce the new painkiller. Under these circumstances, by what method should this decision be made?

In ⟨Ex. 7.1.1⟩, let μ(minutes) denote the mean time until the analgesic effect of the new painkiller appears. The hypothesis that it is the same as the existing painkiller can be expressed as $H_0: \mu = 30$, and the hypothesis that it is faster than the existing product can be expressed as $H_1: \mu < 30$. In this example, the main question of interest is whether there is clear evidence that the analgesic effect of the new painkiller occurs more quickly than that of the existing product.

Accordingly, if an observed outcome that is unlikely to occur under $H_0: \mu = 30$ but likely to occur under $H_1: \mu < 30$ is obtained, we would reject the hypothesis $H_0: \mu = 30$ and accept $H_1: \mu < 30$ as the alternative. For example, the probability of observing the outcome $\bar{x} \le 29$ under $H_0: \mu = 30$ is [10§]

$$P(\bar{x} \le 29) = P\left(\frac{\bar{x} - \mu}{5/\sqrt{100}} \le -2 \mid H_0\right) \fallingdotseq 0.023$$

10§] 여기에서는 새로운 진통제의 진통효과가 나타나는 시간의 분포가 정규분포이고 평균은 μ, 표준편차는 5분이라고 가정하고 계산한 값이며, 표준편차가 미지의 모수인 경우는 7장 2절에서 다룰다.

which corresponds to a result that would occur only about two or three times out of 100 experiments. Therefore, if such a result is observed, it may be regarded as clear evidence against $H_0: \mu = 30$.

In this example, $H_0: \mu = 30$ is called the null hypothesis, $H_1: \mu < 30$ the alternative hypothesis, and the regions $\bar{x} \le 29$ and $\bar{x} > 29$ are called the rejection region and acceptance region of the null hypothesis H_0, respectively.

In general, the hypothesis that is to be rejected when clear counterevidence is obtained from the observations is called the null hypothesis, and the hypothesis proposed as an alternative in such a case is called the alternative hypothesis. The procedure of deciding whether to reject or accept the null hypothesis based on a sample is referred to as a test. Thus, a test is characterized by the rejection region of the null hypothesis, and a statistic used to define this rejection region — such as the sample mean in the above example — is called a test statistic.

In such tests, errors in decision-making may occur whether the null hypothesis is rejected or accepted, and these can be classified as follows:

Table. 7.1.1. Two types of errors in a test.

		가정의 사실 여부	
		H_0이 사실	H_1이 사실
검정 결과	H_0을 채택	옳은 결정	TYPE II ERROR
	H_0을 기각(H_1을 채택)	TYPE I ERROR	옳은 결정

* Type I error: H_0 is true, H_0 rejected
* Type II error: H_1 is true, H_1 rejected ($\Leftrightarrow$ H_0 accepted)

It is desireable to make the probabilities of such errors as small as possible. In particular, since the null hypothesis is the hypothesis to be rejected when clear counterevident is present, it is customary to use a test in which the probability of committing an error associated with this decision, namely a Type I error - does not exceed a prespecified small value. The value α, which represents the maximum allowable probability of committing a Type I error, is called a significance level, and commonly used significance level include $\alpha = 0.10$, $\alpha = 0.05$, and $\alpha = 0.01$

Definition: Significance level:

maximum allowable

The probability of rejecting H_0 (or accepting H_1) given that H_0 is true is called significance level. i.e., A test such that:

$$P(\text{Reject} \mid H_0) \le \alpha$$

is called a test with significance level α.

Ex. 7.1.2.

In ⟨Ex. 7.1.1⟩, the null hypothesis and alternative hypothesis are:

$$H_0 : \mu = 30 \quad \text{vs.} \quad H_1 : \mu < 30$$

$\bar{X} \sim N(\mu, 5^2/100) \Leftrightarrow Z = \frac{\bar{X}-\mu}{5/10} \sim N(0,1)$. The test with rejection region "$\bar{X} \le 29$" satisfies

$$P(\bar{X} \le 29 \mid H_0) = P\left(\frac{\bar{X}-\mu}{5/10} \le -2\right) \fallingdotseq 0.023 \le 0.05$$

Thus, it is a significance level $\alpha = 0.05$ test. Meanwhile if the Significance level is given as $\alpha = 0.05$,

$$P\left(\frac{\bar{X}-\mu}{5/10} \le -z_{0.05} \mid H_0\right) = 0.05 \le \alpha$$

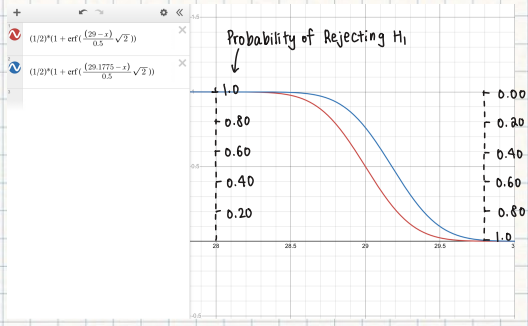
Therefore $\bar{X} \le 30 - 1.645 \times 5/\sqrt{100} = 29.1775$ is a test with significance level $\alpha = 0.05$

Ex 7.1.3.

In ⟨Ex. 7.1.1⟩, the null and alternative hypotheses are

$$H_0 : \mu = 30 \quad \text{vs.} \quad H_1 : \mu < 30 \qquad \text{and} \qquad \bar{X} \sim N(\mu, 5^2/100), \text{ i.e.,}$$

$Z = \frac{\bar{X}-\mu}{5/10} \sim N(0,1)$. Thus, if the true mean is μ, we can express the probabilities of rejecting the null hypothesis by the^ tests from ⟨Ex. 7.1.2⟩ using the standard normal distribution's cdf, Φ. Visualization is in ⟨Image. 7.1.1⟩.



$$P_\mu(\bar{X} \le 29) = P_\mu\left(\frac{\bar{X}-\mu}{5/\sqrt{100}} \le \frac{29-\mu}{5/\sqrt{100}}\right) = \Phi(58-2\mu)$$

$$P_\mu(\bar{X} \le 29.1775) = P_\mu\left(\frac{\bar{X}-\mu}{5/\sqrt{100}} \le \frac{29.1775-\mu}{5/\sqrt{100}}\right) = \Phi(58.355-2\mu)$$

Type Ⅱ error is the probability of accepting H_0 (or rejecting H_1) given that $H_1 : \mu < 30$ is true. As one can see, the test with the rejection region "$\bar{X} \le 30 - 1.645 \times 5/\sqrt{100} = 29.1775$" has a smaller probability of Type Ⅱ error. Meaning, within the permissible range, increasing the size of the rejection region decreases the probability of a Type Ⅱ error.

As can be seen from ⟨Ex. 7.1.3⟩, once the significance level is specified, the rejection region is chosen to be as large as possible within the allowable limit so as to reduce the probability of an incorrect acceptance. Accordingly, for a given significance level α, the rejection region is determined so that

$$P(\text{reject } H_0 \mid H_0) = \alpha$$

and is chosen to be as large as possible subject to this constraint.

Therefore, in ⟨Ex. 7.1.1⟩, in order to test

$$H_0 : \mu = 30 \quad \text{vs.} \quad H_1 : \mu < 30$$

at significance level α, the rejection region is chosen to satisfy:

$$P\left(\frac{\bar{X}-\mu}{5/\sqrt{100}} \le -z_\alpha \mid H_0\right) = \alpha$$

Hence, the rejection region at significance level α is given by

$$\bar{X} \le 30 - z_\alpha \cdot \frac{5}{\sqrt{100}}$$

Ex. 7.1.4.

Here, we are testing $H_0 : \theta = 0.1$ vs. $H_1 : \theta < 0.1$ using the ^mean, $\bar{X}$, of 100 random samples in a Poisson distribution with mean θ. At significance level $\alpha = 0.05$, we will investigate whether we can formulate a rejection region such that

$$P(\text{reject } H_0 \mid H_0) = \alpha$$

In this case, observing a small sample mean is a proof for the alternative hypothesis and a counter-evidence for the null hypothesis. Thus, the form/shape of the test will be: $\bar{X} \le c$. Meanwhile, from the property of Poisson distribution we know that

Probability of Type Ⅱ error

$$100\bar{X} = X_1 + \cdots + X_{100} \sim \text{Poisson}(100\theta)$$

Using the sum of samples, we can reformulate our test as:

$$X_1 + \cdots + X_{100} \le K \quad (K = 100c)$$

Under the null hypothesis, $X_1 + \cdots + X_{100} \sim \text{Poisson}(10)$. Hence the probability of the rejection region is:

$$P(\text{Reject } H_0 \mid H_0) = P_{0.1}(X_1 + \cdots + X_{100} \le K) = P(\text{Poisson}(10) \le K)$$

From Appendix Ⅲ's Poisson distribution's cdf, the probability of rejecting the null hypothesis for different values of K is:

K 가능	0	1	2	3	4	5	…
$P_{0.1}(X_1 + \cdots + X_{100} \le K)$	0.000	0.000	0.002	0.010	0.029	0.067	…

From the table above, we know that no value of k would maximize the rejection region "$X_1+\cdots+X_{100}\le k$" such that its significance level equals to $\alpha=0.05$. However, we can think of probabilistically rejecting the null hypothesis at the boundary in which the probability of the rejection region exceeds 0.05. This type of test is called the randomized test.

$$\begin{cases} \text{If } X_1+\cdots+X_{100}\le 4, & \text{reject } H_0 \\ \text{If } X_1+\cdots+X_{100}=5, & \text{reject } H_0 \text{ with probability } \gamma \\ \text{If } X_1+\cdots+X_{100}\ge 6, & \text{accept } H_0 \end{cases}$$

The rejection probability of this randomized test is:

$$P_{0.1}(X_1+\cdots+X_n\le 4)+\gamma P(X_1+\cdots+X_{100}=5)=0.029+\gamma(0.067-0.029)$$

Thus, we can choose γ so that the rejection probability equals the significance level $\alpha=0.05$:

$$P_{0.1}(X_1+\cdots+X_{100}\le 4)+\gamma P(X_1+\cdots+X_{100}=5)=0.05$$

$$\therefore \gamma=(0.05-0.029)/(0.067-0.029)=21/38$$

Generally, in the case that the population distribution's pdf is one of $f(x;\theta),\theta\in\Omega$, the null hypothesis H_0 and the alternative hypothesis H_1 are:

$$H_0:\theta\in\Omega_0 \quad \text{vs} \quad H_1:\theta\in\Omega_1 \quad (\Omega_0\cap\Omega_1=\phi,\ \Omega_0\cup\Omega_1=\Omega)$$

and the process of testing is judging how clear the evidences or counter-evidences for the null/alternative hypothesis are based on the observations of random samples $X_1,\cdots,X_n$.

In such tests, we set the maximum/upper bound probability of falsely rejecting the null hypothesis as significance level α and set the rejection region/critical region "$(X_1,\cdots,X_n)^t\in C_\alpha$"

$$\text{If } H_0:\theta\in\Omega_0 \text{ is true, } P_\theta\{(X_1,\cdots,X_n)^t\in C_\alpha\}\le\alpha$$

Within the tests with significance level α, it is desirable to minimize the type II error (probability of accepting H_0 when H_1 is true). Thus in order to maximize the rejection/critical region C_α, choose C_α such that

$$\max_{\theta\in\Omega_0} P_\theta\{(X_1,\cdots,X_n)^t\in C_\alpha\}=\alpha$$ and the test becomes:

$$\begin{cases} \text{If } (X_1,\cdots,X_n)^t\in C_\alpha, & \text{reject } H_0 \text{ (accept } H_1) \\ \text{If } (X_1,\cdots,X_n)^t\notin C_\alpha, & \text{accept } H_0 \end{cases}$$

If there is no C_α satisfying $P\{(X_1,\cdots,X_n)^t\in C_\alpha\}$ as in ⟨Ex.7.1.4⟩, we can consider randomized tests. We can express randomized tests using function $\phi(x_1,\cdots,x_n)$ which shows the probability of rejecting the null hypothesis H_0 as the following:

$$\begin{cases} \text{If } \phi(x_1,\cdots,x_n)=1, & \text{reject } H_0 \\ \text{If } 0<\sum\phi(x_1,\cdots,x_n)<1 & \text{reject } H_0 \text{ with probability } \phi(x_1,\cdots,x_n) \\ \text{If } \phi(x_1,\cdots,x_n)=0, & \text{accept } H_0 \end{cases}$$

The rejection probability of such randomized test is given as $E_\theta\phi(X_1,\cdots,X_n)$. $E_\theta[\phi(x_1,\cdots,x_n)]$ is a function of parameter θ and is called the "test function". It is denoted as:

$$\gamma_\phi(\theta)=E_\theta[\phi(X_1,\cdots,X_n)],\ \theta\in\Omega$$

The maximum probability of type I error is called the "size" of test $\phi(x_1,\cdots,x_n)$ and is denoted as:

$$\max_{\theta\in\Omega_0} E_\theta[\phi(x_1,\cdots,x_n)]$$

Tests that satisfy significance level α have sizes lesser than or equal to α:

$$\max_{\theta\in\Omega_0} E_\theta\phi(x_1,\cdots,x_n)\le\alpha$$

Among these tests, it is desirable to choose a test that minimizes the probability of falsely accepting the null hypothesis. Therefore, in order to make the rejection/critical region as wide as possible, we use a test whose size equals the significance level:

$$\max_{\theta\in\Omega_0} E_\theta[\phi(x_1,\cdots,x_n)]=\alpha$$

Ex. 7.1.5.

Let us investigate testing the hypothesis:

$$H_0:\mu\ge 30 \quad \text{vs.} \quad H_1:\mu<30 \quad \text{at significance level } \alpha$$

for mean μ of a normal distribution. Let us suppose that the population distribution is $N(\mu,\sigma^2)$ ($\sigma=5$) and there are 100 iid random samples from this distribution.

Similar to ⟨Ex.7.1.1⟩, smaller sample mean is an evidence for the alternative hypothesis and a counter-evidence for the null hypothesis. Hence, the test will be of the form "$\bar{X}\le c$". In this case, the probability of the rejection/critical region can be expressed using Φ, cdf of standard normal distribution and we see that it is a decreasing function of μ, the population parameter:

$$P_\mu(\bar{X}\le c)=P\left(\frac{\bar{X}-\mu}{5/\sqrt{100}}\le\frac{c-\mu}{5/\sqrt{100}}\right)=\Phi\left(\frac{c-\mu}{5/\sqrt{100}}\right)$$

Hence, when $H_0:\mu\ge 30$ is true, the probability of rejecting H_0 is maximum when $\mu=30$. In order for the maximum type I error to equal α, i.e.,

$$\max_{\mu\ge 30} P_\mu(\bar{X}\le c)=P_{30}(\bar{X}\le c)=\Phi\left(\frac{c-30}{5/\sqrt{100}}\right)=\alpha$$

c must equal to $30-z_\alpha\cdot 5/\sqrt{100}$. From this, when testing the

hypothesis:

$$H_0: \mu \geq 30 \quad \text{vs.} \quad H_1: \mu < 30,$$

Similar to ⟨Ex.7.1.7⟩, the rejection/critical region with significance level α is:

$$\text{"}\ \bar{X} \leq 30 - z_\alpha \cdot 5/\sqrt{100}$$

The rejection/critical region with significance level $\alpha = 0.05$ is

$$\bar{X} \leq 30 - 1.645 \cdot 5/\sqrt{100} = 29.1775$$

The power function (검정력 함수) of this test is given as follows and its graph is equivalent to ⟨Image 7.1.1⟩:

$$\gamma(\mu) = P_\mu(\bar{X} < 29.1775) = P\left(\frac{\bar{X}-\mu}{5/\sqrt{100}} \leq \frac{29.1775-\mu}{5/\sqrt{100}}\right) = \Phi(58.355 - 2\mu)$$

Ex. 7.1.6

In ⟨Ex. 7.1.4⟩ the null/alternative hypotheses for mean θ of Poisson(θ) is:

$$H_0: \theta = 0.1 \quad \text{vs.} \quad H_1: \theta < 0.1$$

randomized

The test with significance level $\alpha = 0.05$ using iid random samples $X_1, \cdots, X_{100}$ is:

$$\phi(x_1, \cdots, x_{100}) = \begin{cases} 1, & x_1 + \cdots + x_{100} \leq 4 \\ 21/38, & x_1 + \cdots + x_{100} = 5 \\ 0, & x_1 + \cdots + x_{100} \geq 6 \end{cases}$$

Meanwhile from sample distribution theory (표본분포 이론) that $X_1 + \cdots + X_{100} \sim$ Poisson(100θ). The power function of this test is the rejection probability shown below, and its graph can be drawn using R. ⟨Image 7.1.2⟩

$$\gamma(\theta) = P_\theta(X_1 + \cdots + X_{100} \leq 4) + \frac{21}{38} P_\theta(X_1 + \cdots + X_{100} = 5)$$

$$= \sum_{x=0}^{4} e^{-100\theta}(100\theta)^x/x! + \frac{21}{38} e^{-100\theta}(100\theta)^5/5!$$



7.2. LIKELIHOOD RATIO TESTS

Suppose that the population distribution is modeled as one of the probability density functions $f(x;\theta), \theta \in \Omega$. We now consider methods for testing hypotheses using a random sample $X_1, \cdots, X_n$ drawn from this population.

The maximum likelihood testing is comparing the maximum likelihood of observations under each hypotheses:

$$H_0: \theta \in \Omega_0 \quad \text{vs.} \quad H_1: \theta \in \Omega_1 \qquad (\Omega_0 \cap \Omega_1 = \emptyset,\ \Omega_0 \cup \Omega_1 = \Omega)$$

Let $L(\theta;x) = \prod_{i=1}^{n} f(x_i;\theta)$ be the likelihood function. We then compare the maximum likelihood under each hypotheses as:

$$\max_{\theta \in \Omega_1} L(\theta;x) \Big/ \max_{\theta \in \Omega_0} L(\theta;x)$$

If this value is large, it is an evidence for the alternative hypothesis and thus we can argue to reject the null hypothesis H_0.

However the maximum likelihood in the entire parameter space, $\max_{\theta \in \Omega} L(\theta;x)$, is the larger likelihood between $\max_{\theta \in \Omega_1} L(\theta;x)$ and $\max_{\theta \in \Omega_0} L(\theta;x)$

Thus,

$$\max_{\theta \in \Omega} L(\theta;x) \Big/ \max_{\theta \in \Omega_0} L(\theta;x) = \max\left\{1, \max_{\theta \in \Omega_1} L(\theta;x) \Big/ \max_{\theta \in \Omega_0} L(\theta;x)\right\}$$

and large $\max_{\theta \in \Omega_1} L(\theta;x) / \max_{\theta \in \Omega_0} L(\theta;x)$ means large $\max_{\theta \in \Omega} L(\theta;x) / \max_{\theta \in \Omega_0} L(\theta;x)$

Hence the rejection/critical region using likelihoods is given as:

$$\max_{\theta \in \Omega} L(\theta;x) \Big/ \max_{\theta \in \Omega_0} L(\theta;x) \geq \lambda$$

and choose λ for a designated value α. Now, the rejection/critical region can also be expressed as the following, using log-likelihoods and the MLEs in each parameter spaces (null/alternative)

$$\ell(\theta;x) = \log L(\theta;x), \quad \hat{\theta}^{\Omega} = \operatorname*{argmax}_{\theta \in \Omega} \ell(\theta;x), \quad \hat{\theta}^{\Omega_0} = \operatorname*{argmax}_{\theta \in \Omega_0} \ell(\theta;x)$$

$$\text{rejection/critical region}: \ 2\left(\ell(\hat{\theta}^{\Omega};x) - \ell(\hat{\theta}^{\Omega_0};x)\right) \geq c$$

Method of Testing using Maximum Likelihoods.

The method of testing

$$H_0: \theta \in \Omega_0 \quad \text{vs.} \quad H_1: \theta \in \Omega_1 \qquad (\Omega_0 \cap \Omega_1 = \emptyset,\ \Omega_0 \cup \Omega_1 = \Omega)$$

at significance level α, using the likelihood as follows is called the "maximum likelihood ratio test" or simply "likelihood ratio test".

1) Shape/form of rejection/critical region:

$$C_\alpha = \left\{x : \max_{\theta \in \Omega} L(\theta;x) \Big/ \max_{\theta \in \Omega_0} L(\theta;x) \geq \lambda\right\}$$

$$= \left\{x : 2\left(\ell(\hat{\theta}^{\Omega};x) - \ell(\hat{\theta}^{\Omega_0};x)\right) \geq c\right\}$$

2) Condition of significance level: $\max_{\theta \in \Omega_0} P_\theta\{(X_1, \cdots, X_n)^t \in C_\alpha\} = \alpha$ (이용)

[09] 필요한 검정에는 이 조건을 위하여 〈예 7.1.4〉 와 같이 랜덤화 검정을 고려한다.

Ex. 7.2.1. Two-Sided test for mean of Normal Distribution.

Find the Likelihood Ratio Test for:

$H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ (μ_0 는 주어진 값)

using random samples $X_1, \cdots, X_n$ $(n \geq 2)$ from a normal distribution $N(\mu, \sigma^2)$ $-\infty < \mu < +\infty$, $0 < \sigma^2 < +\infty$

(Solution)

The log-likelihood and the MLE in each parameter space is:

$$\ell(\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi$$

$$\theta = (\mu, \sigma^2)^t,\ \hat{\theta}^{\Omega} = \left(\bar{x},\ \sum_{i=1}^{n}(x_i-\bar{x})^2/n\right)^t,\ \hat{\theta}^{\Omega_0} = \left(\mu_0,\ \sum_{i=1}^{n}(x_i-\mu_0)^2/n\right)$$

$$\therefore\ 2\{\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}^{\Omega_0})\} = n\log(\hat{\sigma}^{2\Omega_0}/\hat{\sigma}^{2\Omega}) = n\log\left(1 + (\bar{x}-\mu_0)^2/\hat{\sigma}^{2\Omega}\right)$$

Hence the shape of LRT is rejection/critical region is:

$$\left|\frac{\bar{x}-\mu_0}{S/\sqrt{n}}\right| \geq c,\quad S^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)$$

Meanwhile, if $H_0: \mu = \mu_0$ is true, $\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$

Thus the LRT is as follows: Significance level α rejection/critical region

$$: \left|\frac{\bar{x}-\mu_0}{S/\sqrt{n}}\right| \geq t_{\alpha/2}(n-1)$$

This test is called the "two sided t-test".

Ex. 7.2.2. One-Sided test for mean of Normal distribution.

Find the LRT for testing:

$H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$ (μ_0 is given)

at significance level α using random samples $X_1, \cdots, X_n$ from $N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, $0 < \sigma^2 < +\infty$.

(Solution)

$$\hat{\theta}^{\Omega} = (\hat{\mu}^{\Omega}, \hat{\sigma}^{2\Omega}) = \left(\bar{x},\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$$

$$\hat{\theta}^{\Omega_0} = (\hat{\mu}^{\Omega_0}, \hat{\sigma}^{2\Omega_0}) = \left(\hat{\mu}^{\Omega_0},\ \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}^{\Omega_0})^2\right),\ \hat{\mu}^{\Omega_0} = \bar{x} \wedge \mu_0$$

$$\log f(x_i;\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\ell(\theta; x) = \sum \log f(x_i;\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

$$2\{\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}^{\Omega_0})\} = n\log(\hat{\sigma}^{2\Omega_0}/\hat{\sigma}^{2\Omega})$$

$$n\log(\hat{\sigma}^{2\Omega_0}/\hat{\sigma}^{2\Omega}) = n\log\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}\wedge\mu_0)^2 \Big/ \hat{\sigma}^{2\Omega}\right\}$$

$$= n\log\left\{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 + (\bar{x}-\bar{x}\wedge\mu_0)^2\right)/\hat{\sigma}^{2\Omega}\right\}$$

$$= n\log\left\{1 + (\bar{x}-\bar{x}\wedge\mu_0)^2/\hat{\sigma}^{2\Omega}\right\}$$

Test: $(\bar{x}-\bar{x}\wedge\mu_0)^2/\hat{\sigma}^{2\Omega} \geq c$

$$\Leftrightarrow \frac{(\bar{x}-\bar{x}\wedge\mu_0)^2}{\frac{n-1}{n}S^2} \geq c \Leftrightarrow \frac{|\bar{x}-\bar{x}\wedge\mu_0|}{\sqrt{\frac{n-1}{n}}S} \geq c \Leftrightarrow \frac{\bar{x}-\bar{x}\wedge\mu_0}{S/\sqrt{n}} \geq c$$

* $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

The probability of the rejection region is:

$$\gamma(\mu, \sigma^2) = P_{\mu,\sigma^2}\left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}} \geq c\right)$$

and from 〈Theorem 4.2.2〉 we know that

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1),\ V = (n-1)S^2/\sigma^2 \sim \chi^2(n-1) \quad \text{and} \quad Z \overset{*}{\perp\!\!\!\perp} V$$

* independence

Hence the probability of the rejection region is:

$$\gamma(\mu, \sigma^2) = P\left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}} \geq c\right) = P\left(\frac{\bar{x}-\mu+\mu-\mu_0}{\sigma/\sqrt{n}} \Big/ \sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)} \geq c\right)$$

$$= P\left(\frac{Z + (\mu-\mu_0)/(\sigma/\sqrt{n})}{\sqrt{V/(n-1)}} \geq c\right)$$

$$= P\left(Z + (\mu-\mu_0)/(\sigma/\sqrt{n}) \geq c\sqrt{V/(n-1)}\right)$$

$$= P\left((\mu-\mu_0)/(\sigma/\sqrt{n}) \geq c\sqrt{V/(n-1)} - Z\right)$$

Hence, $\gamma(\mu, \sigma^2)$ is an increasing function of $\dfrac{\mu-\mu_0}{\sigma/\sqrt{n}}$ ($\Leftrightarrow$ increasing function of μ. The probability of rejecting H_0, given that H_0 is true, is maximum when $\mu = \mu_0$ (maximizing "size" → maximize critical region → minimize type II error / maximize power)

* (ChatGPT): When the rejection region satisfying $P(\text{reject}\,|\,H_0) = \alpha$ is unique, enlarging the rejection region up to this probability constraint directly maximizes power and hence minimizes type II error.

$$\Rightarrow \max_{(\mu,\sigma^2)\in\Omega} P_{\mu,\sigma^2}\left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}} \geq c\right) = P_{\mu_0,\sigma^2}\left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}} \geq c\right)$$

In addition, when $H_0: \mu = \mu_0$ is true, $\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$

Thus by letting $c = t_\alpha(n-1)$, the test will be size α and have Significance level α. Meaning, the rejection region of the LRT with significance level α $(0 < \alpha < 1)$ is as follows and is called the one sided t test:

$$\frac{\bar{x}-\mu_0}{S/\sqrt{n}} \geq t_\alpha(n-1) \quad : \text{rejection region of significance level } \alpha$$

☆ Ex. 7.2.3. Test for the mean in Exponential distributions.

Find the LRT with Significance level α $(0<\alpha<1)$ for:

$$H_0: \theta=\theta_0 \quad \text{vs.} \quad H_1: \theta\neq\theta_0 \quad (\theta_0 \text{ is given})$$

using random samples $X_1,\cdots,X_n$ from $\text{Exp}(\theta)$

$$f(x;\theta)=\frac{1}{\theta}e^{-x/\theta}I_{(0,\infty)}(x), \quad \log f(x;\theta)=-\log\theta-\frac{x}{\theta}$$

$$\ell(\theta)=\sum_{i=1}^{n}\log f(x_i;\theta)=-n\log\theta-\frac{n\bar{x}}{\theta}$$

$$\hat{\theta}^{\Omega}=\bar{x}, \quad \hat{\theta}^{\Omega_0}=\theta_0$$

$$\therefore \text{LRT}: 2\{\ell(\hat{\theta}^{\Omega})-\ell(\hat{\theta}^{\Omega_0})\}>c \Leftrightarrow 2\left[-n\log\bar{x}-\frac{n\bar{x}}{\bar{x}}+n\log\theta_0+\frac{n\bar{x}}{\theta_0}\right]\geq c$$

$$\Leftrightarrow 2\left\{n\log\left(\frac{\theta_0}{\bar{x}}\right)-n+n\frac{\bar{x}}{\theta_0}\right\}\geq c$$

$$\Leftrightarrow -\log(\bar{x}/\theta_0)+\bar{x}/\theta_0\geq c$$

*notes
$$\begin{cases} -\log(t)\geq -t+1 & \forall t>0 \\ -\log(t)\geq -t+1\geq -t & \forall t>0 \\ -\log(t)+t\geq 0 & \forall t>0 \end{cases}$$

$$\therefore \text{Rejection region}: \begin{cases} \bar{x}/\theta_0\leq C_1 \\ \bar{x}/\theta_0\geq C_2 \end{cases}$$

$$+C_1-\log C_1=C_2-\log C_2$$

$$\text{Exp}(\theta)=\text{Gamma}(1,\theta), \quad E[X]=\theta, \quad \text{Var}(X)=\theta^2 \quad \frac{\bar{X}-\theta}{\theta/\sqrt{n}}\sim N(0,1) \quad \because \text{CLT}$$

$$\sqrt{n}\left(\frac{\bar{X}}{\theta_0}-1\right)=\frac{\sqrt{n}(\bar{X}-\theta_0)}{\theta_0}\sim N(0,1) \text{ under } H_0$$

$$\therefore \text{Rejection region}: \begin{cases} \sqrt{n}(\bar{x}-\theta_0)/\theta_0\leq C_1', & C_1'=\sqrt{n}(C_1-1) \\ \sqrt{n}(\bar{x}-\theta_0)/\theta_0\geq C_2', & C_2'=\sqrt{n}(C_2-1) \end{cases}$$

Your approach - using the CLT to approximate the sample distribution - is called the "Asymptotic LRT". In this exercise, the author implements the "Exact LRT".

$$\text{Exp}(\theta)=\text{Gamma}(1,\theta) \quad \frac{X}{\theta}\sim\text{Gamma}(1,1), \quad \sum\frac{X_i}{\theta}\sim\text{Gamma}(n,1)$$

$$\Rightarrow 2\frac{n\bar{x}}{\theta}\sim\text{Gamma}(n,2)\overset{d}{\equiv}\chi^2(2n)$$

Hence, the rejection region is given as:

$$\text{Rejection region with Significance level } \alpha: \begin{cases} \bar{x}/\theta_0\leq C_1 \text{ or } \bar{x}/\theta_0\geq C_2 \\ C_1-\log C_1=C_2-\log C_2 \\ \int_{2nC_1}^{2nC_2}pdf_{\chi^2(2n)}(y)\,dy=1-\alpha \end{cases}$$

110) 표본크기 n이 클 경우에는 다음 근사식이 성립하는 것이 알려져 있다.

$$C_1\simeq\chi^2_{1-\alpha/2}(2n)/2n\simeq 1-z_{\alpha/2}/\sqrt{n}, \quad C_2\simeq\chi^2_{\alpha/2}(2n)/2n\simeq 1+z_{\alpha/2}/\sqrt{n}$$

기각역의 근사: $\left|\left(\sqrt{n}\left(\bar{x}/\theta_0-1\right)\right)^2\geq z^2_{\alpha/2}=\chi^2_{\alpha}(1)\right.$

☆ Ex. 7.2.4. LRT for comparing the mean of two normal distributions.

Find the LRT with significance level α $(0<\alpha<1)$ for:

$$H_0: \mu_1=\mu_2 \quad \text{vs.} \quad H_1: \mu_1\neq\mu_2$$

using iid random samples $X_{11},\cdots,X_{1n_1}$ from $N(\mu_1,\sigma^2)$ and $X_{21},\cdots,X_{2n_2}$ from $N(\mu_2,\sigma^2)$, $-\infty<\mu_i<+\infty$ $(i=1,2)$, $0<\sigma^2<+\infty$ $(n_1\geq 2, n_2\geq 2)$.

$$\log f(x;\theta)=-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma^2-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}$$

$$\ell(\theta)=\sum_{i=1}^{2}\sum_{j=1}^{n_i}\log f(x_{ij};\theta)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(x_{ij}-\mu_i)^2}{\sigma^2}, \quad (n=n_1+n_2)$$

$$\hat{\mu}_1^{\Omega_0}=\hat{\mu}_2^{\Omega_0}=\frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}x_{ij}, \quad \hat{\mu}_1^{\Omega}=\frac{1}{n_1}\sum_{j=1}^{n_1}x_{1j}, \quad \hat{\mu}_2^{\Omega}=\frac{1}{n_2}\sum_{j=1}^{n_2}x_{2j}$$

$$\partial\ell/\partial\sigma=-\frac{n}{\sigma}+\sum_{i=1}^{2}\sum_{j=1}^{n_i}\frac{(x_{ij}-\mu_i)^2}{\sigma^3}>0$$

$$=\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2>n\sigma^2$$

$$\therefore \hat{\sigma}^{2\Omega}=\frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\hat{\mu}_i^{\Omega})^2, \quad \hat{\sigma}^{2\Omega_0}=\frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\hat{\mu}_i^{\Omega_0})^2$$

$$2\{\ell(\hat{\theta}^{\Omega})-\ell(\hat{\theta}^{\Omega_0})\}=-\frac{n}{2}\log\frac{\hat{\sigma}^{2\Omega}}{\hat{\sigma}^{2\Omega_0}}\geq c$$

$$=\log\frac{\hat{\sigma}^{2\Omega_0}}{\hat{\sigma}^{2\Omega}}\geq c$$

$$\hat{\sigma}^{2\Omega_0}/\hat{\sigma}^{2\Omega}=\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2\Big/\left\{\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2\right\}$$

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=\sum_{j=1}^{n_1}(x_{1j}-\bar{x})^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x})^2$$

$$=\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1+\bar{x}_1-\bar{x})^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2+\bar{x}_2-\bar{x})^2$$

$$=\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2+n_1(\bar{x}_1-\bar{x})^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2+n_2(\bar{x}_2-\bar{x})^2$$

$$\bar{x}_1-\bar{x}=\bar{x}_1-\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\frac{n_2(\bar{x}_1-\bar{x}_2)}{n_1+n_2}$$

$$\bar{x}_2-\bar{x}=\bar{x}_2-\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}=\frac{n_1(\bar{x}_1-\bar{x}_2)}{n_1+n_2}$$

$$\therefore \sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2=\sum_{j=1}^{n_1}(x_{1j}-\bar{x}_1)^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2+\frac{n_1n_2^2(\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^2}+\frac{n_2n_1^2(\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^2}$$

$$=n\cdot\hat{\sigma}^{2\Omega}+n_1n_2(n_1+n_2)^{-2}n(\bar{x}_1-\bar{x}_2)^2$$

$$\therefore \hat{\sigma}^{2\Omega_0}=\frac{1}{n}\sum\sum(x_{ij}-\bar{x})^2=\hat{\sigma}^{2\Omega}+n_1n_2(n_1+n_2)^{-2}(\bar{x}_1-\bar{x}_2)^2$$

$$\hat{\sigma}^{2\Omega_0}/\hat{\sigma}^{2\Omega}=1+n_1n_2(n_1+n_2)^{-2}(\bar{x}_1-\bar{x}_2)^2/\hat{\sigma}^{2\Omega}$$

$$\therefore \text{LRT}: 1+n_1n_2(n_1+n_2)^{-2}(\bar{x}_1-\bar{x}_2)^2/\hat{\sigma}^{2\Omega}\geq c$$

$$\Leftrightarrow \frac{(\bar{x}_1-\bar{x}_2)^2}{(n_1^{-1}+n_2^{-1})S_p^2\cdot(n_1+n_2-2)}\geq c, \quad S_p^2=\frac{1}{n_1+n_2-2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$$

$$\Leftrightarrow \left|\frac{\bar{x}_1-\bar{x}_2}{\sqrt{n_1^{-1}+n_2^{-1}}\cdot S_p}\right|\geq c \qquad n\hat{\sigma}^{2\Omega}=(n_1+n_2)\hat{\sigma}^{2\Omega}=(n_1+n_2-2)S_p^2$$

$$\frac{\bar{x}_1-\bar{x}_2}{\sqrt{n_1^{-1}+n_2^{-1}}S_p}\sim t(n_1+n_2-2) \quad \therefore c=t_{\alpha/2}(n-2) \quad *$$

$$\bar{x}_1\sim N(0,\sigma^2/n_1)$$
$$\bar{x}_2\sim N(0,\sigma^2/n_2)$$
$$\bar{x}_1-\bar{x}_2\sim N(0,(n_1^{-1}+n_2^{-1})\sigma^2)$$

$$\therefore \frac{\bar{x}_1-\bar{x}_2}{(n_1^{-1}+n_2^{-1})\sigma^2}\sim N(0,1), \quad \frac{\bar{x}_1-\bar{x}_2}{(n_1^{-1}+n_2^{-1})\sigma^2}\Big/\sqrt{\frac{(n-2)S_p^2}{\sigma^2}/(n-2)}\sim t(n-2)$$

Ex 7.2.5 Test for comparing the variance between two normal distributions.

Find the LRT with significance level α $(0<\alpha<1)$ for testing:

$$H_0: \sigma_1^2=\sigma_2^2 \quad \text{vs.} \quad H_1: \sigma_1^2\neq\sigma_2^2$$

using random samples $X_{11},\cdots,X_{1n_1}$ from $N(\mu_1,\sigma_1^2)$ and $X_{21},\cdots,X_{2n_2}$ from $N(\mu_2,\sigma_2^2)$ $\quad -\infty<\mu_i<+\infty$, $0<\sigma_i^2<+\infty$ $(i=1,2)$, $(n_1\geq 2, n_2\geq 2)$

$$\ell(\theta)=-\frac{n}{2}\log(2\pi)-\frac{n_1}{2}\log\sigma_1^2-\frac{n_2}{2}\log\sigma_2^2-\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\frac{x_{ij}-\mu_i}{\sigma_i}\right)^2$$

$$\hat{\mu}_i^{\Omega_0}=\hat{\mu}_i^{\Omega}=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}\quad (i=1,2)\ =\bar{x}_i$$

$$\partial\ell/\partial\sigma_i=-\frac{n_i}{\sigma_i}+\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2/\sigma_i^3>0$$

$$\therefore\ \sigma_i^2<\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2$$

$$\therefore\ \hat{\sigma}_i^{2\Omega}=\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-\hat{\mu}_i^{\Omega})^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$$

under H_0, $\ell(\theta;x)=-\frac{n}{2}\log(2\pi)-\frac{n_1+n_2}{2}\log\sigma^2-\frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\frac{x_i-\mu_i}{\sigma}\right)^2$

$$\partial\ell/\partial\sigma=-\frac{n_1+n_2}{\sigma}+\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_i-\mu_i)^2/\sigma^3>0$$

$$\therefore\ \hat{\sigma}^{2\Omega_0}=\frac{1}{n_1+n_2}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_i-\hat{\mu}_i^{\Omega})^2$$

$$\Rightarrow\ (n_1+n_2)\hat{\sigma}^{2\Omega_0}=n_1\hat{\sigma}_1^{2\Omega}+n_2\hat{\sigma}_2^{2\Omega}$$

$$\therefore\ \hat{\sigma}^{2\Omega_0}=(n_1+n_2)^{-1}\left\{n_1\hat{\sigma}_1^{2\Omega}+n_2\hat{\sigma}_2^{2\Omega}\right\}$$

LRT: $2(\ell(\hat{\theta}^{\Omega})-\ell(\hat{\theta}^{\Omega_0}))\geq c$

$$\Rightarrow\ -\frac{n_1}{2}\log(\hat{\sigma}_1^{2\Omega})-\frac{n_2}{2}\log(\hat{\sigma}_2^{2\Omega})+\frac{n}{2}\log(\hat{\sigma}^{2\Omega_0})\geq c$$

$$\Rightarrow\ \frac{n_1}{2}\log(\hat{\sigma}^{2\Omega_0}/\hat{\sigma}_1^{2\Omega})+\frac{n_2}{2}\log(\hat{\sigma}^{2\Omega_0}/\hat{\sigma}_2^{2\Omega})\geq c$$

$$\Rightarrow\ \frac{n_1}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1+n_2\frac{\hat{\sigma}_2^{2\Omega}}{\hat{\sigma}_1^{2\Omega}}\right)\right\}+\frac{n_2}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1\frac{\hat{\sigma}_1^{2\Omega}}{\hat{\sigma}_2^{2\Omega}}+n_2\right)\right\}\geq c$$

let $t=\hat{\sigma}_1^{2\Omega}/\hat{\sigma}_2^{2\Omega}$

$$\text{LRT}=\frac{n_1}{2}\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2t^{-1}\right\}\right\}+\frac{n_2}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1t+n_2\right)\right\}\geq c$$

$n_1+n_2t^{-1}\geq c_1$ or $n_1t+n_2\geq c_2$

$\therefore\ t\geq c'$ or $t\leq c''$: critical region.

$n_1\hat{\sigma}_1^2/\sigma_1^2\sim\chi^2(n_1)$ and $n_2\hat{\sigma}_2^2/\sigma_2^2\sim\chi^2(n_2-1)$

$$\frac{(n_1\hat{\sigma}_1^2/\sigma_1^2)/(n_1-1)}{(n_2\hat{\sigma}_2^2/\sigma_2^2)/(n_2-1)}\sim\frac{V_1/(n_1-1)}{V_2/(n_2-1)},\quad V_1\sim\chi^2(n_1-1),\ V_2\sim\chi^2(n_2-1)$$

$$*\ \ \frac{n_1\hat{\sigma}_1^2}{n_1-1}=S_1^2,\ \frac{n_2\hat{\sigma}_2^2}{n_2-1}=S_2^2\ \therefore\ \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

$$S_i^2=\frac{1}{n_i-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$\therefore\ \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

under H_0, $\sigma_1^2=\sigma_2^2$. Thus $S_1^2/S_2^2\sim F(n_1-1,n_2-1)$ under H_0

LRT: ① $t\geq c_1$ or $t\leq c_2$ $(c_1>c_2)\Leftrightarrow S_1^2/S_2^2\geq d_1$ or $S_1^2/S_2^2\leq d_2$ $(d_2<d_1)$

② $\int_{d_2}^{d_1}pdf_{S_1^2/S_2^2}(x)\,dx=1-\alpha$, ③ : below.

$$\frac{n_1}{2}\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2t^{-1}\right\}\right\}+\frac{n_2}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1t+n_2\right)\right\}\geq c$$

(c_1 & c_2 대응)

$$\therefore\ \frac{n_1}{2}\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2c_1^{-1}\right\}\right\}+\frac{n_2}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1c_1+n_2\right)\right\}$$

$$=$$

$$\frac{n_1}{2}\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2c_2^{-1}\right\}\right\}+\frac{n_2}{2}\log\left\{(n_1+n_2)^{-1}\left(n_1c_2+n_2\right)\right\}$$

$$\Downarrow$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2c_1^{-1}\right\}\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_1c_1+n_2\right)\right\}$$

$$=$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2c_2^{-1}\right\}\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_1c_2+n_2\right)\right\}$$

$$t=\hat{\sigma}_1^2/\hat{\sigma}_2^2\quad\left[n\hat{\sigma}_1^2/(n_1-1)\right]/\left[n_2\hat{\sigma}_2^2/(n_2-1)\right]=S_1^2/S_2^2$$

$$\therefore\ \frac{n_1}{n_2}\cdot\frac{n_2-1}{n_1-1}\cdot t=S_1^2/S_2^2.$$

Here, $\ell(t=c_1)=\ell(t=c_2)\Leftrightarrow\ell(S_1^2/S_2^2=d_1)=\ell(S_1^2/S_2^2=d_2)$ ··· ③

$$\therefore\ d_1=\frac{n_1}{n_2}\cdot\frac{n_2-1}{n_1-1}\cdot c_1,\quad d_2=\frac{n_1}{n_2}\cdot\frac{n_2-1}{n_1-1}\cdot c_2$$

$$c_1=\frac{n_2}{n_1}\cdot\frac{n_1-1}{n_2-1}d_1,\quad c_2=\frac{n_2}{n_1}\cdot\frac{n_1-1}{n_2-1}d_2$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2\frac{n_1}{n_2}\cdot\frac{n_2-1}{n_1-1}\cdot d_1^{-1}\right\}\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_1\frac{n_2}{n_1}\cdot\frac{n_1-1}{n_2-1}d_1+n_2\right)\right\}$$

$$=$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left\{n_1+n_2\frac{n_1}{n_2}\cdot\frac{n_2-1}{n_1-1}\cdot d_2^{-1}\right\}\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_1\frac{n_2}{n_1}\cdot\frac{n_1-1}{n_2-1}d_2+n_2\right)\right\}$$

$$\Downarrow$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left(n_1+n_1\cdot\frac{r_2}{r_1}d_1^{-1}\right)\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_2\frac{r_1}{r_2}d_1+n_2\right)\right\}$$

$$=$$

$$n_1\log\left\{(n_1+n_2)^{-1}\left(n_1+n_1\cdot\frac{r_2}{r_1}\cdot d_2^{-1}\right)\right\}+n_2\log\left\{(n_1+n_2)^{-1}\left(n_2\frac{r_1}{r_2}d_2+n_2\right)\right\}$$

$$(\cdots)$$

we will stop here.

Typically, the LRT would be:

reject H_0 if $F\leq F_{\alpha/2}(n_1-1,n_2-1)$ or $F\geq F_{1-\alpha/2}(n_1-1,n_2-1)$

Read page 310 for textbook approach.

Ex. 7.2.6 Test for mean of Poisson distribution.

Find the LRT with significance level α for testing:

$$H_0: \theta \geq 0.1 \quad \text{vs.} \quad H_1: \theta < 0.1$$

using random samples $X_1, \cdots, X_n$ $(n=100)$ drawn from Poisson(θ), $0 < \theta < +\infty$

$$f(x;\theta) = e^{-\lambda}\lambda^x/x! \qquad \theta=\lambda$$

$$\log f(x;\theta) = -\theta + x\log\theta - \log x!$$
$$\ell(\theta) = -n\theta + \log\theta\, n\bar{x} - \sum_{i=1}^{n}\log x_i!$$

$$\hat{\theta}^{\Omega} = \bar{x}, \quad \hat{\theta}^{\Omega_0} = \bar{x} \vee 0.1$$

$$2(\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}^{\Omega_0})) = -n(\hat{\theta}^{\Omega} - \hat{\theta}^{\Omega_0}) + n\bar{x}\log(\hat{\theta}^{\Omega}/\hat{\theta}^{\Omega_0}) \geq c$$
$$= -n(\bar{x} - \bar{x}\vee 0.1) + n\bar{x}\log(\bar{x}/(\bar{x}\vee 0.1)) \geq c$$

If $\bar{x} \geq 0.1$, $2(\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}^{\Omega_0})) = 0 \Rightarrow$ we always accept H_0

if $\bar{x} < 0.1$, $2(\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}^{\Omega_0})) = -n(\bar{x} - 0.1) + n\bar{x}\log(\bar{x}/0.1) \geq c$
$$= \bar{x} - \bar{x}\log(\bar{x}) \geq c$$

LRT Test: $\bar{x} \leq C$ (Rejection/Critical region)

$\bar{x} \leq C \Leftrightarrow \sum x_i \leq C'$, $\sum x_i \sim$ Poisson(100θ)

under H_0, $\hat{\theta}^{\Omega_0} = 0.1$ $\therefore \sum x_i \sim$ Poisson(10)

x	3	4	5	6	...
$P(X \leq x)$	0.0103	0.0293	0.0671	0.1301	...

If $\sum x_i \leq 4$: reject H_0
If $\sum x_i = 5$: reject with probability γ
If $\sum x_i \geq 6$: Accept H_0

$P(\sum x_i \leq 4) + \gamma P(\sum x_i = 5) = 0.05$
$\Rightarrow 0.0293 + \gamma(0.0671 - 0.0293) = 0.05$
$\therefore \gamma = (0.05 - 0.0293)/(0.0671 - 0.0293) = \frac{23}{42}$

$$\therefore \text{LRT Test} = \begin{cases} \sum x_i \leq 4 : \text{reject } H_0 \\ \sum x_i = 5 : \text{reject } H_0 \text{ with probability } 23/42^{*} \\ \sum x_i \geq 6 : \text{Accept } H_0 \end{cases}$$

* Textbook states 21/38. However, the Poisson probabilities used in the exercise is not precise enough (two decimal places). Hence 23/42 is more accurate (three decimal place poisson probabilities)

7.3. APPROXIMATION OF LIKELIHOOD RATIO TEST

We can approximate the distribution of test statistics in LRT by using the fact that the distribution of MLE is asymptotically normal. Let the log-likelihood function and the mean log-likelihood function of n random samples $X_1, \cdots, X_n$ be:

$$\ell(\theta) = \sum_{i=1}^{n}\log f(x_i;\theta), \quad \bar{\ell}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\log f(x_i;\theta)$$

If Ω is an open set (열린집합) in $\mathbb{R}^K$, let us look into the limiting distribution of the LRT test statistic.

As we have shown in chapter 6, when sample size is large, we can approximate the estimation error of MLE using the information number and the mean score function:

$$\sqrt{n}(\hat{\theta}_n^{MLE} - \theta) \simeq [I(\theta)]^{-1}\sqrt{n}\,\dot{\bar{\ell}}(\theta)$$

Here, the information number and the mean score function are:

$$I(\theta) = \mathrm{Var}_\theta\left\{\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right\} = E_\theta\left\{-\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right\}$$

$$\dot{\bar{\ell}}(\theta) = \left(\frac{\partial}{\partial\theta_1}\bar{\ell}(\theta), \cdots, \frac{\partial}{\partial\theta_K}\bar{\ell}(\theta)\right)^t$$

Given the null hypothesis $H_0: \theta = \theta_0$, the LRT test statistic is:

$$2\left\{\ell(\hat{\theta}_n^{MLE}) - \ell(\theta_0)\right\}$$

For simplicity MLE will be denoted as $\hat{\theta}_n$. Let us look into the limiting distribution of the LRT test statistic.

If we expand the likelihood equation around $\hat{\theta}_n$ using Taylor's theorem,

$$\ell(\theta_0) = \ell(\hat{\theta}_n) + \dot{\ell}(\hat{\theta}_n)^t(\theta_0 - \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^t\ddot{\ell}(\theta_n^*)(\theta_0 - \hat{\theta}_n)/2,$$

$$\exists\,\theta_n^* : |\theta_n^* - \theta_0| \leq |\hat{\theta}_n - \theta_0|$$

Under suitable conditions, $\hat{\theta}_n$ is the root of the likelihood equation, $\dot{\ell}_n(\theta) = 0$, and since $\hat{\theta}_n$ is a consistent estimator, if $H_0: \theta = \theta_0$ is true, $\hat{\theta}_n \simeq \theta_0$ and therefore $\theta_n^* \simeq \theta_0$ (θ_n^* approaches θ_0). Therefore, by the LLN, the following approximation holds.

$$-\frac{1}{n}\ddot{\ell}(\theta_n^*) \simeq -\ddot{\bar{\ell}}(\theta_0) \simeq E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right] = I(\theta_0)$$

Thus, if $H_0: \theta = \theta_0$ is true,

$$\ell(\theta_0) \simeq \ell(\hat{\theta}_n) + \dot{\ell}(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n) - n(\theta_0 - \hat{\theta}_n)^t\frac{1}{n}\ddot{\ell}(\theta_n^*)(\theta_0 - \hat{\theta}_n)/2$$
$$\simeq \ell(\hat{\theta}_n) + 0 - \sqrt{n}(\theta_0 - \hat{\theta}_n)^t I(\theta)\sqrt{n}(\theta_0 - \hat{\theta}_n)/2$$

Meanwhile, under suitable conditions, if $H_0: \theta = \theta_0$ is true

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow[n\to\infty]{d} N(0, I(\theta)^{-1})$$

Hence, by ⟨Theorem 5.3.2⟩ and ⟨Theorem 4.4.5⟩, if $H_0: \theta = \theta_0$ is true, the following holds:

$$2\left\{\ell(\hat{\theta}_n) - \ell(\theta_0)\right\} \simeq \sqrt{n}(\hat{\theta}_n - \theta_0)^t I(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{d} \chi^2(k)$$

⟨Theorem 7.3.1. The limiting distribution of LRT test statistic in simple hypothesis testing⟩

if conditions [R0) ~ [R7] from ⟨Theorem 6.4.4⟩ are satisfied in a probability model whose pdf is $f(x;\theta), \theta \in \Omega \subseteq \mathbb{R}^k$, the following approximation holds for the LRT test statistic in the hypothesis test:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta \neq \theta_0$$

(a) If $H_0: \theta = \theta_0$ is true,

$$2\left\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\theta_0)\right\} \xrightarrow[n\to\infty]{d} \chi^2(k)$$

Thus, when the sample size is large, the asymptotic rejection region having significance level α is:

$$2\left\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\theta_0)\right\} \geq \chi^2_{\alpha}(k)$$

(b) If $H_0: \theta = \theta_0$ is true

$$2\left\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\theta_0)\right\} = n(\hat{\theta}_n^{\Omega} - \theta_0)^t I(\theta_0)(\hat{\theta}_n^{\Omega} - \theta_0) + r_n, \; r_n \xrightarrow[n\to\infty]{P_{\theta_0}} 0$$

Thus, when the sample size is large, the asymptotic rejection region having significance level α is:

$$n(\hat{\theta}_n^{\Omega} - \theta_0)^t I(\theta_0)(\hat{\theta}_n^{\Omega} - \theta_0) \geq \chi^2(k)$$

(c) If $H_0: \theta = \theta_0$ is true

$$2\left\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\theta_0)\right\} = n\bar{\ell}(\theta_0)^t \left[I(\theta_0)\right]^{-1}\bar{\ell}(\theta_0) + r_n, \; r_n \xrightarrow[n\to\infty]{P_{\theta_0}} 0$$

Thus, when the sample size is large, the asymptotic rejection region having significance level α is:

$$n\bar{\ell}(\theta_n)^t \left[I(\theta_0)\right]^{-1}\bar{\ell}(\theta_0) \geq \chi^2_{\alpha}(k)$$

(proof) The rigorous proofs of these asymptotic distributions require a rigorous treatment of the remainder term. Such treatments are similar to the ones shown in ⟨Theorem 6.4.4.⟩. Thus, the proofs will be left out.

The test statistics in ⟨Theorem 7.3.1⟩ (b) and (c) are called the "Wald's test statistic" and "Rao's test statistic" respectively.

$$W_n(\theta_0) = n(\hat{\theta}_n^{\Omega} - \theta_0)^t I(\theta_0)(\hat{\theta}_n^{\Omega} - \theta_0)$$

$$R_n(\theta_0) = n\bar{\ell}_n(\theta_0)^t \left[I(\theta_0)\right]^{-1}\bar{\ell}(\theta_0)$$

Ex. 7.3.1. Approximation of Test for mean of exponential distribution

The LRT for the simple hypothesis test $H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta_1$ using random samples $X_1, \cdots, X_n$ derived from $Exp(\theta)$ was found in ⟨Ex. 7.2.3⟩

Meanwhile, $\ell(\theta) = -n\log\theta - n\bar{x}/\theta$, $\bar{\ell}_n(\theta) = -\frac{1}{\theta} + \frac{\bar{x}}{\theta^2}$, $\hat{\theta}^{\Omega} = \bar{x}$, $I(\theta) = 1/\theta^2$

Thus, the LRT test statistic and its asymptotic test statistics are:

(a) LRT test statistic: $2\left\{\ell(\hat{\theta}^{\Omega}) - \ell(\hat{\theta}_0)\right\} = 2n\left\{\bar{x}/\theta_0 - 1 - \log(\bar{x}/\theta_0)\right\}$

(b) Wald's test statistic: $n(\hat{\theta}^{\Omega_0} - \theta_0)^t I(\theta)(\hat{\theta}^{\Omega_0} - \theta_0)$
$= n(\bar{x} - \theta_0)^2\theta_0^{-2}$

(c) Rao's test statistic: $n\bar{\ell}(\theta_0)^t I(\theta)^{-1}\bar{\ell}(\theta_0) = n\left[-\frac{1}{\theta_0} + \frac{\bar{x}}{\theta^2}\right]^2\theta_0^2$
$= n(\bar{x} - \theta_0)^2\theta_0^{-2}$

☆☆ Try Solving it.

Ex. 7.3.2. Approximation of tests in Multinomial distribution model

We want to test: $H_0: p = p_0$ vs. $H_1: p \neq p_0$ $(p = (p_1, \cdots, p_k)^t)$ using $Z_1, \cdots, Z_n$, which are the results for n $Multi(1, (p_1, \cdots, p_k)^t)$ $(p_1 + \cdots + p_k = 1, \; p_i > 0)$ trials.

Find Wald's and Rao's test statistics and the asymptotic rejection region having significance level α.

(Solution)

$X_i = \sum_{j=1}^{n} Z_{ij}$ $(i = 1, \cdots, k)$, $\theta = (p_1, \cdots, p_r)^t$ $(r = k-1)$, $p_* = \theta_1 + \cdots + \theta_r = 1 - p_k$

If we use the above to express the likelihood function and its partial derivatives,

$$\ell(\theta) = \sum_{i=1}^{r} x_i \log\theta_i + x_k \log(1 - \theta_*), \quad \dot{\ell}(\theta) = \left(x_i/\theta_i - x_k/(1-\theta_*)\right)_{1 \leq i \leq r}$$

From above, if we find the MLE and the information number matrix $I(\theta)$,

$$\hat{\theta}^{\Omega} = (x_1/n, \cdots, x_r/n)^t, \quad \hat{p}_i^{\Omega} = x_i/n \;\; (i = 1, \cdots, k)$$

$$-\ddot{\ell}(\theta) = \mathrm{diag}(x_i/\theta_i^2) + (x_k/(1-\theta_*))^2 \mathbb{1}\mathbb{1}^t, \quad I(\theta) = \mathrm{diag}(1/\theta_i) + p_k^{-1}\mathbb{1}\mathbb{1}^t$$

Hence,

$$I(\theta)(\hat{\theta}^{\Omega} - \theta) = \left[\mathrm{diag}(1/\theta_i) + p_k^{-1}\mathbb{1}\mathbb{1}^t\right]\left(x_i/n - \theta_i\right)_{1 \leq i \leq r}$$

$$= \underbrace{(\hat{\theta}_i^{\Omega}/\theta_i)}_{①} + \underbrace{\left(\sum_1^r x_i/p_k\right)}_{②}{}_{1 \leq i \leq r} - \mathbb{1} - \underbrace{\left(\sum_1^r \theta_i/p_k\right)}_{③}{}_{1 \leq i \leq r}$$

$$= (\hat{\theta}_i^{\Omega}/\theta_i)_{1 \leq i \leq r} \quad \mathbb{1}(1 - \hat{p}_k^{\Omega})/p_k - \mathbb{1} - \mathbb{1}(1-p_k)/p_k$$

$$= \left(\frac{x_i - np_i}{np_i}\right)_{1 \leq i \leq r} - \left(\frac{x_k - np_k}{np_k}\right) \cdot \mathbb{1}$$

$$= \bar{\ell}(\theta)$$

$$\therefore\ n(\hat{\theta}^{\Omega}-\theta)^t I(\theta)(\hat{\theta}^{\Omega}-\theta) = \sum_{i=1}^{K}\frac{(x_i-np_i)^2}{np_i} = n\,\bar{l}(\theta)^t\left[I(\theta)\right]^{-1}\bar{l}(\theta)$$

Wald's — Rao's

Therefore, the LRT test statistic is as follows and under $H_0: p=p_0$, the asymptotic distribution of the test statistic is $\chi^2(K-1)$. If the test statistic is greater than or equal to $\chi^2_\alpha(K-1)$, its test has significance level α.

(a) LRT test statistic: $2\left\{l(\hat{\theta}^{\Omega}) - l(\theta_0)\right\} = 2\sum_{i=1}^{K} X_i \log(X_i/np_{0i})$

(b) Wald's test statistic: $W_n(\theta_0): \sum_{i=1}^{K}(X_i-np_{0i})^2/(np_{0i})$

(c) Rao's test statistic: $R_n(\theta_0): \sum_{i=1}^{K}(X_i-np_{0i})^2/(np_{0i})$

Let $X_i = \sum_{j=1}^{n} Z_{ij}$

likelihood: $\prod_{j=1}^{n} f(z_{1j},\cdots z_{Kj};\theta) = \prod_{j=1}^{n} p_1^{z_{1j}}\cdots p_K^{z_{Kj}}$

$$l(p;z) = \sum_{j=1}^{n} z_{1j}\log p_1 + \cdots + \sum z_{Kj}\log p_K$$

$$= x_1\log p_1 + \cdots + x_K \log p_K$$

To eliminate dependency, let $p_K = 1-\sum_{i=1}^{r} p_i = 1-p_\cdot$, $r = k-1$ and $x_K = n-\sum_i^r x_i = n-x_\cdot$

$$l(\theta) = x_1\log p_1 + \cdots x_r \log p_r + (n-x_\cdot)\log(1-p_\cdot)$$

$$\partial l/\partial p_i = \frac{x_i}{p_i} - \frac{(n-x_\cdot)}{1-p_\cdot} = \frac{x_i}{p_i} - \frac{x_K}{p_K}$$

$$\partial^2 l/\partial p_i\partial p_j = \frac{n-x_\cdot}{(1-p_\cdot)^2} = \frac{x_K}{p_K^2}$$

$$\partial^2 l/\partial p_i^2 = -\frac{x_i}{p_i^2} + \frac{n-x_\cdot}{(1-p_\cdot)^2} = -\frac{x_i}{p_i^2} + \frac{x_K}{p_K^2}$$

$$\therefore\ \ddot{l}(\theta) = \mathrm{diag}\left(-\frac{x_i}{p_i^2}\right) + \mathbb{1}\mathbb{1}^T\frac{x_K}{p_K}$$

$$I(\theta) = E_\theta[-\ddot{l}(\theta)] = \frac{1}{n}\left\{\mathrm{diag}\left(\frac{np_i}{p_i^2}\right) - \mathbb{1}\mathbb{1}^T\frac{np_K}{p_K^2}\right\}$$

$$= \mathrm{diag}\left(\frac{1}{p_i}\right) - \mathbb{1}\mathbb{1}^T\frac{1}{p_K}$$

$$\sqrt{n}(\hat{\theta}^{\Omega}-\theta_0)^T I(\theta)\sqrt{n}(\hat{\theta}^{\Omega}-\theta_0)$$

$$= n\left[(\hat{\theta}^{\Omega}-\theta_0)^T\mathrm{diag}\left(\frac{1}{p_i}\right)(\hat{\theta}^{\Omega}-\theta) - \frac{1}{p_K}(\hat{\theta}^{\Omega}-\theta_0)^T\mathbb{1}\mathbb{1}^T(\hat{\theta}^{\Omega}-\theta_0)\right]$$

$$= n\left[\sum_{i=1}^{r}\frac{\left(\frac{x_i}{n}-p_{0i}\right)^2}{p_i} - \frac{1}{p_K}\left(\sum\frac{x_i}{n}-p_{0i}\right)^2\right]$$

$$= \sum_{i=1}^{r}\frac{(x_i-np_{0i})^2}{np_i} - \frac{n}{p_K}\frac{\left\{\sum(x_i-np_{0i})\right\}^2}{n^2}$$

$$= \sum_{i=1}^{r}\frac{(x_i-np_{0i})^2}{np_i} + \frac{\left\{(n-x_\cdot)-(n-np_\cdot)\right\}^2}{np_K}$$

$$= \sum_{i=1}^{r}\frac{(x_i-np_{0i})^2}{np_i} + \frac{(x_K-np_K)^2}{np_K}$$

$$= \sum_{i=1}^{K}\frac{(x_i-np_{0i})^2}{np_i} = \sqrt{n}\,\bar{l}(\theta_0)\,I(\theta)^{-1}\sqrt{n}\,\bar{l}(\theta_0)$$

Similar to 〈Ex. 7.2.1〉, even though the parameter is a vector, the null hypothesis may only concern the mean. (eg. In 〈Ex. 7.2.1.〉 the test compares the mean between two normal distributions which has two parameters - μ and σ^2 - but the null hypothesis is: $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$). Approximation for LRT test statistic is possible for such cases also.

*증명이 복잡

〈Theorem 7.3.2. The limiting distribution of LRT test statistic in general hypothesis test〉

If

1. parameter space is an open-set in $\mathbb{R}^K$
2. pdf is one of $f(x;\theta)$, $\theta \in \Omega \subseteq \mathbb{R}^K$
3. Conditions (R0) - (R7) in 〈Theorem 6.4.4〉 are satisfied
4. Let ξ_1 be a component of parameter $\theta = (\xi_1^t, \eta^t)^t$. The parameter space $\Omega_0 = \{(\xi_{10}^t, \eta^t)^t \in \Omega : \eta \in \mathbb{R}^{K_0}\}$, that corresponds to the null hypothesis $H_0: \xi_1 = \xi_{10}$, is an open-set in $\mathbb{R}^{K_0}$

The following holds for the LRT test statistic in the hypothesis test:

$$H_0: \xi_1 = \xi_{10} \quad \text{vs.} \quad H_1: \xi_1 \neq \xi_{10}$$

(a) If $H_0: \xi_1 = \xi_{10}$ is true,

$$2\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\hat{\theta}_n^{\Omega_0})\} \xrightarrow[n\to\infty]{d} \chi^2(K-K_0)$$

When sample size is large, the rejection region with significance level α is:

$$2\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\hat{\theta}_n^{\Omega_0})\} \geq \chi_\alpha^2(K-K_0)$$

(b) If $H_0: \xi_1 = \xi_{10}$ is true and $\theta_0 = (\xi_{10}^t, \eta_0^t)^t \in \Omega_0$,

$$2\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\hat{\theta}_n^{\Omega_0})\} = n(\hat{\theta}_n^{\Omega} - \hat{\theta}_n^{\Omega_0})^t I(\theta)(\hat{\theta}_n^{\Omega} - \hat{\theta}_n^{\Omega_0}) + r_n, \quad r_n \xrightarrow[n\to\infty]{P_{\xi_0,\eta}} 0$$

(c) If $H_0: \xi_1 = \xi_{10}$ is true and $\theta_0 = (\xi_{10}^t, \eta_0^t)^t \in \Omega_0$ [111]

$$2\{\ell(\hat{\theta}_n^{\Omega}) - \ell(\hat{\theta}_n^{\Omega_0})\} = n\bar{l}(\hat{\theta}_n^{\Omega_0})^t [I(\theta_0)]^{-1} \bar{l}(\hat{\theta}_n^{\Omega_0}) + r_n, \quad r_n \xrightarrow[n\to\infty]{P_{\xi_0,\eta}} 0$$

(proof) Such approximation equations require the rigorous treatment of the remainder terms which we have see in 〈Theorem 6.4.4〉 Hence, we will introduce the sketch of the proof here.

In order to separate the log-likelihood function and its partial derivatives component-wise, we will use the following notation:

$$\dot{\ell}(\theta) = \begin{pmatrix} \dot{\ell}_1(\theta) \\ \dot{\ell}_2(\theta) \end{pmatrix} = \begin{pmatrix} \partial\ell(\theta)/\partial\xi \\ \partial\ell(\theta)/\partial\eta \end{pmatrix}, \quad I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix}$$

and for simplicity, the MLE of parameters in the entire parameter space (Ω) vs. parameter space under the null hypothesis, which are $\hat{\theta}_n^{\Omega}$ and $\hat{\theta}_n^{\Omega_0}$, respectively, will be denoted as $\hat{\theta}_n$ and $\hat{\theta}_n^{\circ}$.

When the null hypothesis $H_0: \xi_1 = \xi_{10}$ is true, let the corresponding parameter be denoted by $\theta_0 = (\xi_{10}^t, \eta_0^t) \in \Omega_0$. Then, as in the proof of 〈Theorem 7.3.2〉, the following approximation holds.

$$\sqrt{n}(\hat{\theta}_n - \theta) \simeq [I(\theta_0)]^{-1} \sqrt{n}\,\bar{l}(\theta_0)$$

$$2\{\ell(\hat{\theta}_n) - \ell(\theta_0)\} \simeq \sqrt{n}\,\bar{l}(\theta_0)^t [I(\theta_0)]^{-1} \sqrt{n}\,\bar{l}(\theta_0) \quad \cdots ①$$

In the same way,

$$\sqrt{n}(\hat{\eta}_n^{\circ} - \eta_0) \simeq [I_{22}(\theta_0)]^{-1} \sqrt{n}\,\bar{l}_2(\theta_0)$$

$$\ell(\theta_0) \simeq \ell(\hat{\theta}_n^{\circ}) - \sqrt{n}(\eta_0 - \hat{\eta}_n^{\circ})^t I_{22}(\theta_0) \sqrt{n}(\eta_0 - \hat{\eta}_n^{\circ})/2$$

$$2\{\ell(\hat{\theta}_n^{\circ}) - \ell(\theta_0)\} \simeq \sqrt{n}\,\bar{l}_2(\theta_0)^t [I_{22}(\theta_0)]^{-1} \sqrt{n}\,\bar{l}_2(\theta_0) \quad \cdots ②$$

$\theta_0 = (\xi_0^t, \eta_0^t)^t$, $\hat{\theta}_n^{\circ} = (\xi_{10}^t, \hat{\eta}_n^{\circ t})^t$

$\therefore \ell(\theta_0) - \ell(\hat{\theta}_n^{\circ})$

$= \ell(\xi_{10}) + \ell(\eta_0) - \{\ell(\xi_{10}) + \ell(\hat{\eta}_n^{\circ})\}$

$= \ell(\eta_0) - \ell(\hat{\eta}_n^{\circ})$

$= \sqrt{n}\,\bar{l}_2(\theta_0)^t I_{22}(\theta_0)^{-1} \sqrt{n}\,\bar{l}_2(\theta_0)/2$

$$\therefore 2\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^{\circ})\} \simeq \sqrt{n}\,\bar{l}(\theta_0)^t [I(\theta_0)]^{-1} \sqrt{n}\,\bar{l}(\theta_0) - \sqrt{n}\,\bar{l}_2(\theta_0)^t [I_{22}(\theta_0)]^{-1} \sqrt{n}\,\bar{l}_2(\theta_0)$$

↳ $2\{\ell(\hat{\theta}_n) - \ell(\theta_0) + \ell(\theta_0) - \ell(\hat{\theta}_n^{\circ})\}$

$= \underbrace{2\{\ell(\hat{\theta}_n) - \ell(\theta_0)\}}_{①} - \underbrace{2\{\ell(\hat{\theta}_n^{\circ}) - \ell(\theta_0)\}}_{②}$

교재에는 $\sqrt{n}$ 곱함

However, it is known that if $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, [112]

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11\cdot 2}^{-1} & -\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot 2}^{-1} & \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\cdot 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{22}^{-1} \end{bmatrix}$$

Apply this to $[I(\theta)]^{-1}$

$$\sqrt{n}\,\bar{l}(\theta_0)^t [I(\theta_0)]^{-1} \sqrt{n}\,\bar{l}(\theta_0) - \sqrt{n}\,\bar{l}_2(\theta_0)^t [I_{22}(\theta_0)]^{-1} \sqrt{n}\,\bar{l}_2(\theta_0)$$

$$= \sqrt{n}\begin{pmatrix} \bar{l}_1(\theta_1) & \bar{l}_2(\theta_2) \end{pmatrix} [I(\theta_0)]^{-1} \sqrt{n} \begin{pmatrix} \bar{l}_1(\theta_1) \\ \bar{l}_2(\theta_3) \end{pmatrix} - \sqrt{n}\,\bar{l}_2(\theta_0)^t [I_{22}(\theta_0)]^{-1} \sqrt{n}\,\bar{l}_2(\theta_0)$$

$$= \sqrt{n}\begin{pmatrix} \bar{l}_1(\theta_1) & \bar{l}_2(\theta_2) \end{pmatrix} \begin{bmatrix} I_{11\cdot 2}^{-1} & -I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} \\ -I_{22}^{-1} I_{21} I_{11\cdot 2}^{-1} & I_{22}^{-1} I_{21} I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1} \end{bmatrix} \begin{pmatrix} \bar{l}_1(\theta_0) \\ \bar{l}_2(\theta_0) \end{pmatrix}$$

* notice how $+\Sigma_{22}^{-1}$ term is eliminated

$$= \sqrt{n}\,\bar{l}(\theta_0)^t \begin{bmatrix} I_{K_1} \\ -I_{22}^{-1} I_{21} \end{bmatrix} \begin{bmatrix} I_{11\cdot 2}^{-1} \end{bmatrix} \begin{bmatrix} I_{K_1} & -I_{12} I_{22}^{-1} \end{bmatrix} \sqrt{n}\,\bar{l}(\theta_0) \qquad K_1 = \dim(\xi_0)$$

$$= \sqrt{n}\, z_n^t [I_{11\cdot 2}(\theta_0)]^{-1} \sqrt{n}\, z_n, \quad z_n = \bar{l}_1(\theta_0) - I_{12}(\theta_0) I_{22}(\theta_0)^{-1} \bar{l}_2(\theta_0)$$

111) (b), (c)에서 θ_0는 미지의 η_0가 포함되어 있으므로 이를 $\hat{\theta}_n^{\Omega_0}$로 추정한 검정통계량을 사용한다.

112) 부분행렬로 블럭된 행렬에 관한 이러한 사실은 부록 II에 증명되어 있다.

Meanwhile, from the CLT,

$$\sqrt{n}\,Z_n \xrightarrow[n\to\infty]{d} N_{K_1}(0, I_{11.2}(\theta_0)),\quad K_1 = K - K_0$$

* CLT: $\sqrt{n}\left(\frac{1}{n}\sum_i^n T(X_i) - E[T(X)]\right) \to N(0, Var(T(X)))$

$$T(X) = \left[\frac{\partial}{\partial\theta_0}\log f(x;\theta) \;\; \frac{\partial}{\partial\eta_0}\log f(x;\theta)\right]^t = \frac{\partial}{\partial\theta}\log f(x;\theta)$$

$$E[T(X)] = E\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right] = E\left[\frac{f'(x;\theta)}{f(x;\theta)}\right] = \frac{\partial}{\partial\theta}\int_{x\in S_X} f(x;\theta)\,dx = \frac{\partial}{\partial\theta}1 = 0$$

$$Var(T(X)) = Var\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) = I(\theta)$$

Let $g(t_1, t_2) = t_1 - I_{12}(\theta_0) I_{22}^{-1}(\theta_0) t_2$, $\dot{g}(t_1, t_2) = \left[I_{K_1} \;\; -I_{12}(\theta_0) I_{22}^{-1}(\theta_0) I_{K_0}\right]^t$

$$\sqrt{n}(Z_n) = \sqrt{n}\left(g\left(\frac{1}{n}\sum T(X)\right) - g(E[T(X)])\right) \xrightarrow[n\to\infty]{d} \dot{g}(E[T(X)])\,N(0, I(\theta))$$

$$\overset{d}{=} N(0,\ \dot{g}(E[T(X)])^t I(\theta) \dot{g}(E[T(X)])) \overset{d}{=} N(0, I_{11.2})$$

$$\therefore 2\left\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^*)\right\} \simeq \sqrt{n}\, Z_n^t \left[I_{11.2}(\theta_0)\right]^{-1} \sqrt{n}\, Z_n$$

Hence (a) follows.

Meanwhile, if we utilize $I(\theta_0)(\hat{\theta}_n - \theta_0) \simeq \bar{l}(\theta_0)$, $\hat{\eta}_n^\circ - \eta_0 \simeq \left[I_{22}(\theta_0)\right]^{-1}\bar{l}_2(\theta_0)$

$$\begin{bmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{bmatrix}\begin{bmatrix} \hat{\xi}_n - \xi_0 \\ \hat{\eta}_n - \eta_0 \end{bmatrix} = \begin{bmatrix} \bar{l}_1(\theta_0) \\ \bar{l}_2(\theta_0) \end{bmatrix}$$

$$\therefore \bar{l}_2(\theta_0) = I_{21}(\theta)(\hat{\xi}_n - \xi_0) + I_{22}(\theta)(\hat{\eta}_n - \eta_0)$$

$$\Leftrightarrow \hat{\eta}_n - \eta_0 = \left[I_{22}(\theta)\right]^{-1}\left\{\bar{l}_2(\theta_2) - I_{21}(\theta)(\hat{\xi}_n - \xi_0)\right\}$$

$$\therefore \hat{\eta}_n - \hat{\eta}_n^\circ = (\hat{\eta}_n - \eta_0) - (\hat{\eta}_n^\circ - \eta_0)$$

$$= \left[I_{22}(\theta)\right]^{-1}\left\{\bar{l}_2(\theta_2) - I_{21}(\theta)(\hat{\xi}_n - \xi_0)\right\} - \left[I_{22}(\theta_0)\right]^{-1}\bar{l}_2(\theta_0)$$

$$= -\left[I_{22}(\theta)\right]^{-1} I_{21}(\theta)(\hat{\xi}_n - \xi_0) \quad \cdots ①$$

$$I_{11.2}(\theta_0)(\hat{\xi}_n - \xi_0) \simeq \left(I_{11}(\theta_0) - I_{12}(\theta_0) I_{22}(\theta_0)^{-1} I_{21}(\theta_0)\right)(\hat{\xi}_n - \xi_0)$$

$$= I_{11}(\theta_0)\cdot(\hat{\xi}_n - \xi_0) - \underbrace{I_{12}(\theta_0) I_{22}(\theta_0)^{-1} I_{21}(\theta_0)(\hat{\xi}_n - \xi_0)}_{= ①}$$

$$= I_{11}(\theta_0)(\hat{\xi}_n - \xi_0) - I_{12}(\theta_0)\cdot(\hat{\eta}_n - \hat{\eta}_n^\circ)$$

$$= \bar{l}(\theta_0) - I_{12}(\theta_0)(\hat{\eta}_n - \eta_0 + \eta_0 - \hat{\eta}_n^\circ)$$

$$= \bar{l}(\theta_0) - I_{12}(\theta_0) I_{22}^{-1}(\theta_0)\underbrace{I_{22}(\theta_0)(\hat{\eta}_n - \eta_0)}_{\text{consistent}} + I_{12}(\theta_0) I_{22}^{-1}(\theta_0) I_{22}(\theta_0)(\hat{\eta}_n - \eta_0)$$

$$\simeq \bar{l}(\theta_0) - I_{12}(\theta_0) I_{22}^{-1}(\theta_0)\,\bar{l}_2(\theta_0)$$

$\therefore (\hat{\eta}_n - \eta_0) \xrightarrow[n\to\infty]{} 0$

$$\therefore \underbrace{I_{11.2}(\theta_0)(\hat{\xi}_n - \xi_0)}_{= Z_n} \simeq \bar{l}_1(\theta_0) - I_{12}(\theta_0) I_{22}^{-1}(\theta_0)\,\bar{l}_2(\theta_0) \quad \cdots ②$$

$$\therefore (\hat{\theta}_n - \hat{\theta}_n^\circ) = \left((\hat{\xi}_n - \xi_0)^t,\ (-I_{22}(\theta_0)^{-1} I_{21}(\theta_0)(\hat{\xi}_n - \xi_0))^t\right)^t \leftarrow \hat{\theta}_n - \hat{\theta}_n^\circ = \begin{bmatrix} \hat{\xi}_n - \xi_0 \\ \hat{\eta}_n - \hat{\eta}_n^\circ \end{bmatrix}$$

$$\therefore n(\hat{\theta}_n - \hat{\theta}_n^\circ)^t I(\theta_0)(\hat{\theta}_n - \hat{\theta}_n^\circ) \simeq Z_n^t\left[I_{11.2}(\theta_0)\right]^{-1} Z_n \simeq 2\left\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^\circ)\right\}$$

Hence, we know that the approximation in (b) is true.

(c) can be proved in a similar manner.

Steps of proof

(a) ① Find $2\left\{\ell(\hat{\theta}_n) - \ell(\theta_0)\right\}$ (Full parameter space Ω) $\simeq \sqrt{n}\,\bar{l}(\theta_0)^t\left[I(\theta_0)\right]^{-1}\sqrt{n}\,\bar{l}(\theta_0)$

② Find $2\left\{\ell(\hat{\theta}_n^\circ) - \ell(\theta_0)\right\}$ (Null Hypothesis, parameter space Ω_0) $\simeq \sqrt{n}\,\bar{l}_2(\theta_0)^t\left[I_{22}(\theta_0)\right]^{-1}\sqrt{n}\,\bar{l}_2(\theta_0)$

③ Find $2\left\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^\circ)\right\}$ using ① & ②

④ Use $\left[I(\theta)\right]^{-1} = \begin{bmatrix} \Sigma_{11.2}^{-1} & -\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1} & \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{22}^{-1} \end{bmatrix}$

to express ② as $\sqrt{n}\,Z_n^t\left[I_{11.2}\right]^{-1}\sqrt{n}\,Z_n$

$Z_n = \bar{l}_1(\theta_0) - I_{12}(\theta_0) I_{22}(\theta_0)^{-1}\bar{l}_2(\theta_0)$

⑤ Use the CLT to show that $\sqrt{n}\,Z_n \to N(0, I_{11.2}(\theta_0))$

⑥ $\therefore 2\left\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^\circ)\right\} \simeq \sqrt{n}\,Z_n^t I_{11.2}(\theta_0)^{-1}\sqrt{n}\,Z_n$

$$= \left[\sqrt{n}\left(I_{11.2}(\theta_0)^{-1/2}\right)Z_n\right]^t\left[\sqrt{n}\left(I_{11.2}(\theta_0)^{-1/2}\right)Z_n\right]$$

$= z^t z$, $z \sim N(0, I_{K-K_0})$, $\theta^\circ \in \mathbb{R}^K$, $\eta \in \mathbb{R}^{K_1}$

$\sim \chi^2(K - K_0)$

(b)

① Use $\sqrt{n}(\hat{\theta}_n - \theta_0) \simeq \sqrt{n}\, I(\theta_0)^{-1}\bar{l}(\theta_0)$ … (1)

$\sqrt{n}(\hat{\theta}_n^\circ - \theta_0) \simeq \sqrt{n}\, I_{22}(\theta_0)^{-1}\bar{l}_2(\theta_0)$ … (2)

$= \hat{\eta}_n^\circ - \eta_0$

② : ① − (1) : $I(\theta_0)(\hat{\theta}_n - \theta_0) = \bar{l}(\theta_0)$

$$\begin{bmatrix} 0 & I_{K_0} \end{bmatrix} I(\theta_0)(\hat{\theta}_n - \theta_0) = \begin{bmatrix} 0 & I_{K_0} \end{bmatrix}\bar{l}(\theta_0)$$

$$\Rightarrow I_{21}(\theta_0)(\hat{\xi}_n - \xi_0) + I_{22}(\theta_0)\underbrace{(\hat{\eta}_n - \eta_0)}_{*} = \bar{l}_2(\theta_0)$$

rearrange equation w.r.t (*)

$$(\hat{\eta}_n - \eta_0) = I_{22}^{-1}\left[\bar{l}_2(\theta_0) - I_{21}(\hat{\xi}_n - \xi_0)\right]$$

$$(\hat{\eta}_n - \hat{\eta}_n^\circ) = (\hat{\eta}_n - \eta_0) - (\hat{\eta}_n^\circ - \eta_0)$$

$$= -I_{22}^{-1} I_{21}(\hat{\xi}_n - \xi_0)$$ ← Memorize (If you want to write the proof with ease)

③ $(\hat{\theta}_n - \hat{\theta}_n^\circ)^t = (\hat{\xi}_n - \hat{\xi}_0,\ \hat{\eta}_n - \hat{\eta}_n^\circ)^t$

$$= (\hat{\xi}_n - \hat{\xi}_0,\ -I_{22}^{-1} I_{21}(\hat{\xi}_n - \xi_0))^t$$

④ Show that $I_{11.2}(\theta_0)(\hat{\xi}_n - \xi_0) = \bar{l}_1(\theta_0) - I_{12} I_{22}^{-1}\bar{l}_2(\theta_0) = Z_n$

⑤ $n(\hat{\theta}_n - \hat{\theta}_n^\circ)^t I(\theta_0)(\hat{\theta}_n - \hat{\theta}_n^\circ) = (\hat{\xi}_n - \xi_0)^t I_{11.2}(\hat{\xi}_n - \xi_0)$

$$= Z_n^t\left[I_{11.2}\right]^{-1} Z_n^t$$

$$= 2\left\{\ell(\hat{\theta}_n) - \ell(\hat{\theta}_n^\circ)\right\} \sim \chi^2(K - K_0)$$

From the approximations in ⟨Theorem 7.3.2⟩ (b) and (c), we can think of test statistics:

$$W_n(\hat{\theta}_n^0) = n(\hat{\theta}_n - \hat{\theta}_n^0)^t I(\hat{\theta}_n^0)(\hat{\theta}_n - \hat{\theta}_n^0), \quad R_n(\hat{\theta}_n^0) = n\bar{\ell}(\hat{\theta}_n^0)^t [I(\hat{\theta}_n^0)]^{-1}\bar{\ell}(\hat{\theta}_n^0)$$

These are called Wald's and Rao's test statistic respectively and we see that the Wald's/Rao's test statistics are essentially θ_0 substituted by $\hat{\theta}_n^0$ in the test statistics

$$W_n(\theta_0) = n(\hat{\theta}_n - \theta_0)^t I(\theta_0)(\hat{\theta}_n - \theta_0), \quad R_n(\theta_0) = n\bar{\ell}(\theta_0)^t [I(\theta_0)]^{-1}\bar{\ell}(\theta_0)$$

which tests the hypothesis $H_0: \theta = \theta_0$. Also given that the information number matrix $I(\theta)$ is continuous for θ, the approximation of these test statistics are the same as that of ⟨Theorem 7.3.2⟩

Meanwhile, if the null hypothesis is given as:

$$H_0: g_1(\theta) = 0, \cdots, g_{K_1}(\theta) = 0 \quad (K_1 = K - K_0)$$

where $g_1(\theta), \cdots, g_{K_1}(\theta)$ are differentiable,

$$\xi_1 = g_1(\theta), \cdots, \xi_K = g_{K_1}(\theta), \quad \eta_1 = g_{K_1+1}(\theta), \cdots, \eta_{K_0} = g_K(\theta) \quad (K_1 + K_0 = k)$$

We can transform the parameters using differentiable functions as above and test for

$$H_0: \xi = 0$$

Thus, even for these kind of H_0, we can perform ⟨Theorem 7.3.2⟩'s approximate tests.

---

* 교수님께서 수업중에는 다루시지 않으셨지만 복습해 볼 것.

★ Ex. 7.3.3. Independence Test in Contingency Tables.

Let X_{ij} $(i = 1, \cdots, r), (j = 1, \cdots, c)$ in on $(r \times c)$ contigency table for a bivariate classification be the observations from

$$X = (X_{ij}) \sim \text{Multi}(n, (p_{ij})), \quad \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1, \quad p_{ij} > 0 \quad (i = 1, \cdots, r), (j = 1, \cdots, c)$$

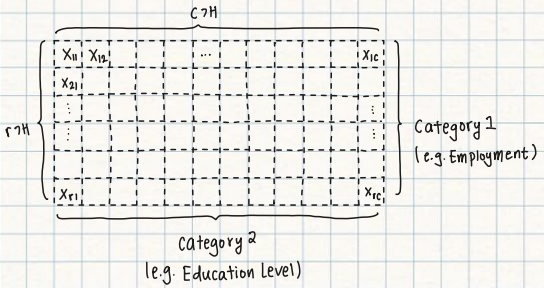
The null hypothesis for independence is:

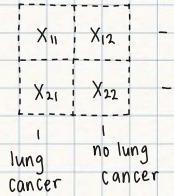
$$H_0: p_{ij} = (p_{i\cdot})(p_{\cdot j}) \quad (i = 1, \cdots, r)\ (j = 1, \cdots, c) \quad \left(p_{i\cdot} = \sum_{j=1}^{c} p_{ij},\ p_{\cdot j} = \sum_{i=1}^{r} p_{ij} \right)$$

H_1: not H_0

When sample size is large, find the LRT whose asymptotic significance level is α.

⟨Table for better understanding⟩





H_0: Smoking & lung cancer are independent

vs

H_1: Smoking & lung cancer are dependent.

* For binary variables (such as smoking - lung cancer example), dependence infers correlation.

(Solution)

$$\ell(p) = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij} \log p_{ij} + (\text{constant}), \quad p_{ij} > 0, \quad \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1 \quad \begin{matrix}(i = 1, \cdots, r)\\ (j = 1, \cdots, c)\end{matrix}$$

Let $(x_{i\cdot}) = \sum_{j=1}^{c} x_{ij}$, $(x_{\cdot j}) = \sum_{i=1}^{r} x_{ij}$, $p^0 = (p_{i\cdot}^0, p_{\cdot j}^0)$.

Under H_0, the log-likelihood function is:

$$\ell(p^\circ) = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log p_{i\cdot}^\circ p_{\cdot j}^\circ + (\text{constant})$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log p_{i\cdot}^\circ + x_{ij}\log p_{\cdot j}^\circ + (\text{constant})$$

$$= \sum_{i=1}^{r} x_{i\cdot}\log p_{i\cdot}^\circ + \sum_{j=1}^{c} x_{\cdot j}\log p_{\cdot j}^\circ + (\text{constant}),\quad p_{i\cdot}^\circ > 0,\ \sum_{i=1}^{r} p_{i\cdot}^\circ = 1,\ p_{\cdot j}^\circ > 0,\ \sum_{j=1}^{c} p_{\cdot j}^\circ = 1$$

These are all log-likelihood functions of multinomial distributions or the sum of the functions that we have covered in ⟨Ex. 6.3.6⟩ and ⟨Ex. 7.3.2⟩.

Hence the MLE in Ω (entire parameter space) and Ω° (parameter space under H_0) are:

$$\hat{p}_{ij} = x_{ij}/n,\quad \hat{p}_{ij}^\circ = \hat{p}_{i\cdot}^\circ \times \hat{p}_{\cdot j}^\circ = (x_{i\cdot}/n)(x_{\cdot j}/n)$$

Thus, the LRT test statistic is:

$$2\left\{\ell(\hat{p}) - \ell(\hat{p}^\circ)\right\} = 2\sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log\left[x_{ij}/(n\hat{p}_{i\cdot}^\circ\hat{p}_{\cdot j}^\circ)\right]$$

Meanwhile, if the parameter $p^\circ = (p_{i\cdot}^\circ, p_{\cdot j}^\circ)$ has a designated value, i.e.,

$$H_0(p^\circ):\ p_{ij} = p_{i\cdot}^\circ p_{\cdot j}^\circ \quad (i=1,\cdots,r)\ (j=1,\cdots,c)$$

, as we have shown in ⟨Ex. 7.3.2⟩, The Wald's and Rao's test statistics are:

$$W_n(p^\circ) = R_n(p^\circ) = \sum_{i=1}^{r}\sum_{j=1}^{c} (x_{ij} - np_{i\cdot}^\circ p_{\cdot j}^\circ)^2/(np_{i\cdot}^\circ p_{\cdot j}^\circ)$$

Thus, as we have shown in ⟨Theorem 7.3.2⟩, if we substitute the unknown parameters $p^\circ = (p_{i\cdot}^\circ, p_{\cdot j}^\circ)$ with estimates under H_0, that is $\hat{p}^\circ$, we get the following test-statistic.

$$W_n(\hat{p}^\circ) = R_n(\hat{p}^\circ) = \sum_{i=1}^{r}\sum_{j=1}^{c} (x_{ij} - n\hat{p}_{i\cdot}^\circ\hat{p}_{\cdot j}^\circ)^2/(n\hat{p}_{i\cdot}^\circ\hat{p}_{\cdot j}^\circ)$$

In this case, the Ω and Ω_0 are respectively:

$$\Omega = \left\{(p_{ij}) : \sum_{i=1}^{r}\sum_{j=1}^{c} p_{ij} = 1,\ p_{ij} > 0\right\}$$

$$\Omega_0 = \left\{(p_{i\cdot}, p_{\cdot j}) : \sum_{i=1}^{r} p_{i\cdot} = 1, \sum_{j=1}^{c} p_{\cdot j} = 1,\ p_{i\cdot} > 0,\ p_{\cdot j} > 0\right\}$$

and their dimensions are:

$$K = rc - 1,\quad K_0 = (r-1) + (c-1)$$

Hence, the distribution of the test statistic under H_0 is:

$$\chi^2(K - K_0) = \chi^2(rc - 1 - (r-1) - (c-1)) = \chi^2(rc - r - c + 1)$$
$$= \chi^2((r-1)(c-1))$$

Generally the rejection region using Wald's/Rao's test statistic is:

$$\sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij}^\circ)^2}{E_{ij}^\circ} \geq \chi^2((r-1)(c-1);\alpha)$$

$O_{ij} = x_{ij}$: observation, E_{ij}° : Expected value under H_0

This test is called the chi-squared test in a contingency table.

Up to this point, we have considered only the use of maximum likelihood based on random samples, that is, observations that are iid. However, estimation and testing using the maximum likelihood method can also be carried out based on observations arising from more general settings. In particular, when the observations are independent, results on limiting distributions in ⟨Theorem 7.3.1⟩ and ⟨Theorem 7.3.2⟩ can be obtained under suitable regularity conditions.

In such cases, the definition of the likelihood and the log-likelihood remain the same, and the information matrix is defined as:

$$I_n(\theta) = E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\sum_{i=1}^{n}\log f(x_i;\theta)\right],\quad \bar{I}_n(\theta) = I_n(\theta)/n$$

and under H_0, the Wald's/Rao's test statistics are:

$$W_n(\theta_0) = n(\hat{\theta}_n^\Omega - \theta_0)^t\bar{I}_n(\theta_0)(\hat{\theta}_n^\Omega - \theta_0),\quad R_n(\theta_0) = n\bar{\ell}(\theta_0)^t\left[\bar{I}_n(\theta_0)\right]^{-1}\bar{\ell}(\theta_0)$$

In simple hypothesis testing, substitute θ_0 with $\hat{\theta}_n^\circ$ and perform approximate tests as ⟨Theorem 7.3.2⟩.

★★ Ex 7.3.4. Test of homogeneity in a contingency table.

$$\begin{cases} X_i = (X_{i1},\cdots,X_{ic})^t \sim \text{Multi}(n_i, p_i),\ p_i = (p_{i1},\cdots,p_{ic})^t,\ i=1,\cdots,r \\ X_1,\cdots,X_r \text{ are independent.}\end{cases}$$

In a model representing observations from mutually independent multinomial distributions, a chi-square approximation test based on the following test statistic can also be used to test the null hypothesis of homogeneity, that is, to test the hypothesis

$$H_0: p_1 = \cdots = p_r \quad \text{vs.} \quad H_1: p_1,\cdots,p_r \text{ are different.}$$

In this case, the degrees of freedom are given by

$$K - K_0 = r(c-1) - (c-1) = (r-1)(c-1)$$

(a) LRT Test statistic: $2\left\{\ell(\hat{p}) - \ell(\hat{p}^\circ)\right\} = 2\sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log\left[x_{ij}/(n_i\hat{p}_{ij}^\circ)\right]$

$$\hat{p}_{ij}^\circ = \cdots = \hat{p}_{cj}^\circ = \sum_{i=1}^{r} x_{ij}/n,\ j=1,\cdots,c$$

$n = n_1 + \cdots + n_r$

(b) Wald's & Rao's Test statistic:

$$W_n(\hat{p}^\circ) = R_n(\hat{p}^\circ) = \sum_{i=1}^{r}\sum_{j=1}^{c} (x_{ij} - n_i\hat{p}_{ij}^\circ)^2/(n_i\hat{p}_{ij}^\circ)$$

Here, the approximation supposes that n_i $(i=1,\cdots,r)$ is increasing and to summarize the calculation process:

$$n_i \longrightarrow \infty,\quad \frac{n_i}{n_1+\cdots+n_r} = \gamma_i\ (0 < \gamma_i < 1),\ n = n_1+\cdots+n_r$$

(1) $2(\ell(\hat{\theta}) - \ell(\hat{\theta}^\circ))$: $\ell(\theta) = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}\log p_{ij} + \text{constant}$

$$\hat{p}_i = x_i/n_i,\quad \hat{p}_1^\circ = \cdots = \hat{p}_r^\circ = x_{\cdot j}/n = (x_1+\cdots+x_r)/n$$

$$\therefore\ 2\left\{\ell(\hat{\theta}) - \ell(\hat{\theta}^\circ)\right\} = 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_i\hat{p}_{ij}\log(\hat{p}_{ij}/\hat{p}_{ij}^\circ)$$

(2) Approximation of $2(\ell(\hat\theta)-\ell(\hat\theta^\circ))$:

$$2(\ell(\hat\theta)-\ell(\hat\theta^\circ)) = 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_i\hat p_{ij}^\circ\left(\hat p_{ij}/\hat p_{ij}^\circ\right)\log\left(\hat p_{ij}/\hat p_{ij}^\circ\right)$$

$$\simeq 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_i\hat p_{ij}^\circ\left\{\left(\hat p_{ij}/\hat p_{ij}^\circ-1\right)+\frac{1}{2}\left(\hat p_{ij}/\hat p_{ij}^\circ-1\right)^2\right\} \quad (*)$$

* $t\log t \simeq (t-1)+\frac{1}{2}(t-1)^2$

$$\Rightarrow 2\sum\sum\left(n_i\hat p_{ij}-n_i\hat p_{ij}^\circ\right)+\sum\sum n_i\left(\hat p_{ij}-\hat p_{ij}^\circ\right)^2/\hat p_{ij}^\circ$$

$$\Rightarrow 0 \qquad +\sum\sum\frac{(n_i\hat p_{ij}-n_i\hat p_{ij}^\circ)^2}{n_i\hat p_{ij}^\circ} = \sum\sum\frac{(O_{ij}-\hat E_{ij}^\circ)^2}{\hat E_{ij}^\circ}$$

$$\therefore 2(\ell(\hat\theta)-\ell(\hat\theta^\circ)) \simeq \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-\hat E_{ij}^\circ)^2}{\hat E_{ij}^\circ},\quad O_{ij}=x_{ij},\ \hat E_{ij}^\circ = n_i\hat p_{ij}^\circ = n_i\sum_{i=1}^{r}x_{ij}/n$$

(3) Distribution approximation under $H_0: p_1=\cdots=p_r\equiv p^\circ$

under H_0, $\hat p_i^\circ \simeq p^\circ$. Thus,

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-\hat E_{ij}^\circ)^2}{\hat E_{ij}^\circ} = \sum_{i=1}^{r}\sum_{j=1}^{c} n_i(\hat p_{ij}-\hat p_{ij}^\circ)^2/\hat p_{ij}^\circ$$

$$\simeq \sum_{i=1}^{r}\sum_{j=1}^{c} n_i(\hat p_{ij}-\hat p_{ij}^\circ)^2/p_j^\circ$$

$$= \sum_{i=1}^{r} n_i(\hat\theta_i-\hat\theta_i^\circ)^t\Sigma_0^{-1}(\hat\theta_i-\hat\theta_i^\circ)$$

Here, $\hat\theta_i$, $\hat\theta_i^\circ$ denotes the vector formed by the first $(c-1)$ components in $\hat p_i$ and $\hat p_i^\circ$ respectively.

$\Sigma = \mathrm{diag}(\theta_j^\circ)-\theta^\circ\cdot(\theta^\circ)^t$, $\theta^\circ=(p_1^\circ,\cdots,p_{c-1}^\circ)^t$

Now, let $Z_i=\Sigma_0^{-1/2}\sqrt{n_i}(\hat\theta_i-\theta^\circ)$. Then z_i are mutually independent and under H_0,

$$Z_i \sim N_d(0,I),\quad d=c-1$$

$$\hat\theta_i-\theta^\circ = \sum_{l=1}^{r} n_l(\hat\theta_l-\theta_l^\circ)/n = \sum_{l=1}^{r}\sqrt{n}\,\Sigma_0^{1/2}Z_l/n$$

$$\therefore \Sigma_0^{-1/2}(\hat\theta_i^\circ-\theta^\circ) = \sum_{l=1}^{r} w_lZ_l/w^tw,\quad w_l=\sqrt{n_l},\ w=(w_1,\cdots,w_r)^t$$

Hence, if $A\otimes B=(a_{ij}B)$, use the idempotent matrix

$$I_r\otimes I_d - w(w^tw)^{-1}w^t\otimes I_d$$

to show: $\sum_{i=1}^{r} n_i(\hat\theta_i-\hat\theta_i^\circ)^t\Sigma_0^{-1}(\hat\theta_i-\hat\theta_i^\circ)$
$= (Z_1^t,\cdots,Z_r^t)(I_r\otimes I_d - w(w^tw)^{-1}w^t\otimes I_d)(Z_1^t,\cdots,Z_r^t)^t$

However, $(Z_1^t,\cdots,Z_r^t)^t \sim N_{rd}(0,I)$

Thus, the distribution of this quadratic form is asymptotically:

$$\chi^2(rd-d)=\chi^2((r-1)(c-1)) \quad \square.$$