2.5.3 Generalized likelihood ratio tests

When a UMP test does not exist, we usually use a generalized likelihood ratio test to verify $H_0 : \boldsymbol{\vartheta} \in \Theta^\star$ against $H_1 : \boldsymbol{\vartheta} \in \Theta \backslash \Theta^\star$. It can be used when H_0 is composite, which none of the above methods can.

The generalized likelihood ratio test has rejection region $\mathcal{R} = \{\boldsymbol{y} : \lambda(\boldsymbol{y}) \leq a\}$, where

$$\lambda(\boldsymbol{y}) = \frac{\max_{\boldsymbol{\vartheta} \in \Theta^\star} L(\boldsymbol{\vartheta}|\boldsymbol{y})}{\max_{\boldsymbol{\vartheta} \in \Theta} L(\boldsymbol{\vartheta}|\boldsymbol{y})}$$

is the *generalized likelihood ratio* and a is a constant chosen to give significance level α, that is such that

$$P(\lambda(\boldsymbol{Y}) \leq a | H_0) = \alpha.$$

If we let $\hat{\boldsymbol{\vartheta}}$ denote the maximum likelihood estimate of $\boldsymbol{\vartheta}$ and let $\widehat{\boldsymbol{\vartheta}}_0$ denote the value of $\boldsymbol{\vartheta}$ which maximises the likelihood over all values of $\boldsymbol{\vartheta}$ in $\Theta^\star$, then we may write

$$\lambda(\boldsymbol{y}) = \frac{L(\widehat{\boldsymbol{\vartheta}}_0|\boldsymbol{y})}{L(\widehat{\boldsymbol{\vartheta}}|\boldsymbol{y})}.$$

The quantity $\widehat{\boldsymbol{\vartheta}}_0$ is called the **restricted maximum likelihood estimate** of $\boldsymbol{\vartheta}$ under H_0.

Example 2.37. Suppose that $Y_i \underset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Then the likelihood is

$$L(\mu, \sigma^2|\boldsymbol{y}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}.$$

The maximum likelihood estimate of $\boldsymbol{\vartheta} = (\mu, \sigma^2)^{\mathrm{T}}$ is $\widehat{\boldsymbol{\vartheta}} = (\widehat{\mu}, \widehat{\sigma^2})^{\mathrm{T}}$, where $\widehat{\mu} = \overline{y}$ and $\widehat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2/n$.

Similarly, the restricted maximum likelihood estimate of $\boldsymbol{\vartheta} = (\mu, \sigma^2)^{\mathrm{T}}$ under H_0 is $\widehat{\boldsymbol{\vartheta}}_0 = (\widehat{\mu}_0, \widehat{\sigma^2}_0)^{\mathrm{T}}$, where $\widehat{\mu}_0 = \mu_0$ and $\widehat{\sigma}_0^2 = \sum_{i=1}^{n}(y_i - \mu_0)^2/n$.

Thus, the generalized likelihood ratio is

$$\begin{aligned}
\lambda(\boldsymbol{y}) = \frac{L(\mu_0, \hat{\sigma}_0^2|\boldsymbol{y})}{L(\hat{\mu}, \hat{\sigma}^2|\boldsymbol{y})} &= \frac{(2\pi\hat{\sigma}_0^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (y_i - \mu_0)^2\right\}}{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n (y_i - \overline{y})^2\right\}} \\
&= \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{\frac{n}{2}} \exp\left\{\frac{n\sum_{i=1}^n (y_i - \overline{y})^2}{2\sum_{i=1}^n (y_i - \overline{y})^2} - \frac{n\sum_{i=1}^n (y_i - \mu_0)^2}{2\sum_{i=1}^n (y_i - \mu_0)^2}\right\} \\
&= \left\{\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2}\right\}^{\frac{n}{2}}.
\end{aligned}$$

Since the rejection region is $\mathcal{R} = \{\boldsymbol{y} : \lambda(\boldsymbol{y}) \leq a\}$, we reject H_0 if

$$\left\{\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2}\right\}^{\frac{n}{2}} \leq a \Rightarrow \frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2} \leq b,$$

where a and b are constants chosen to give significance level α. Now, we may write

$$\begin{aligned}
\sum_{i=1}^n (y_i - \mu_0)^2 &= \sum_{i=1}^n \{(y_i - \overline{y}) + (\overline{y} - \mu_0)\}^2 \\
&= \sum_{i=1}^n (y_i - \overline{y})^2 + 2(\overline{y} - \mu_0)\sum_{i=1}^n (y_i - \overline{y}) + n(\overline{y} - \mu_0)^2 \\
&= \sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2.
\end{aligned}$$

So we reject H_0 if

$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2} \leq b.$$

Thus, rearranging, we reject H_0 if

$$1 + \frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2} \geq c \Rightarrow \frac{n(\overline{y} - \mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2} \geq d,$$

where c and d are constants chosen to give significance level α, that is we can write

$$\alpha = P(\lambda(\boldsymbol{Y}) \leq a|H_0) = P\left(\frac{n(\overline{Y} - \mu_0)^2}{S^2} \geq d|H_0\right),$$

where $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$.

To get d we need to work out the distribution of $\frac{n(\overline{Y}-\mu_0)^2}{S^2}$ under the null hypothesis. For $Y_i \underset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ we have

$$\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and so, under H_0,

$$\overline{Y} \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right) \quad and \quad \frac{\sqrt{n}\left(\overline{Y}-\mu_0\right)}{\sqrt{\sigma^2}} \sim \mathcal{N}(0,1).$$

This gives

$$\frac{n(\overline{Y}-\mu_0)^2}{\sigma^2} \sim \chi^2_1.$$

Also

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Now we may use the fact that if U and V are independent rvs such that $U \sim \chi^2_{\nu_1}$ and $V \sim \chi^2_{\nu_2}$, then $\frac{U/\nu_1}{V/\nu_2} \sim \mathcal{F}_{\nu_1,\nu_2}$.

Here, $U = \frac{n(\overline{Y}-\mu_0)^2}{\sigma^2}$ and $V = \frac{(n-1)S^2}{\sigma^2}$. Hence, if H_0 is true, we have

$$F = \frac{U/1}{V/(n-1)} = \frac{n(\overline{Y}-\mu_0)^2}{S^2} \sim \mathcal{F}_{1,n-1}.$$

Therefore, we reject H_0 at a significance level α if

$$\frac{n(\overline{y}-\mu_0)^2}{s^2} \geq F_{1,n-1,\alpha},$$

where $F_{1,n-1,\alpha}$ is such that $P(F \geq F_{1,n-1,\alpha}) = \alpha$.

Equivalently, we reject H_0 if

$$\sqrt{\frac{n(\overline{y}-\mu_0)^2}{s^2}} \geq t_{n-1,\frac{\alpha}{2}},$$

that is, if

$$\left|\frac{\overline{y}-\mu_0}{\sqrt{s^2/n}}\right| \geq t_{n-1,\frac{\alpha}{2}}.$$

Of course, this is the usual two-sided t test. □

In fact, many of the standard tests in situations with normal distributions are generalized likelihood ratio tests.

2.5.4 Wilks' theorem

In more complex cases, we have to use the following approximation to find the rejection region. The result is stated without proof.

Theorem 2.9. *Wilks' theorem.*
Assume that the joint distribution of $Y_1, \ldots, Y_n$ *depends on* p *unknown parameters and that, under* H_0*, the joint distribution depends on* p_0 *unknown parameters. Let* $\nu = p - p_0$*. Then, under some regularity conditions, when the null hypothesis is true, the distribution of the statistic* $-2\log\{\lambda(\boldsymbol{Y})\}$ *converges to a* χ^2_ν *distribution as the sample size* $n \to \infty$*, i.e., when* H_0 *is true and* n *is large,*

$$-2\log\{\lambda(\boldsymbol{Y})\} \underset{approx.}{\sim} \chi^2_\nu.$$

Thus, for large n*, the rejection region for a test with approximate significance level* α *is*

$$\mathcal{R} = \{\boldsymbol{y} : -2\log\{\lambda(\boldsymbol{y})\} \geq \chi^2_{\nu,\alpha}\}.$$

□

Example 2.38. Suppose that $Y_i \underset{iid}{\sim} \text{Poisson}(\lambda)$ and consider testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$.

Then we have seen that no UMP test exists in this case. Now, the likelihood is

$$L(\lambda|\boldsymbol{y}) = \frac{\lambda^{\sum_{i=1}^n y_i} e^{-n\lambda}}{\prod_{i=1}^n y_i!}.$$

The maximum likelihood estimate of λ is $\hat{\lambda} = \overline{y}$ and the restricted maximum likelihood estimate of λ under H_0 is $\hat{\lambda}_0 = \lambda_0$. Thus, the generalized likelihood ratio is

$$\begin{aligned}\lambda(\boldsymbol{y}) = \frac{L(\lambda_0|\boldsymbol{y})}{L(\hat{\lambda}|\boldsymbol{y})} &= \frac{\lambda_0^{\sum_{i=1}^n y_i} e^{-n\lambda_0}}{\prod_{i=1}^n y_i!} \frac{\prod_{i=1}^n y_i!}{\overline{y}^{\sum_{i=1}^n y_i} e^{-n\overline{y}}} \\ &= \left(\frac{\lambda_0}{\overline{y}}\right)^{\sum_{i=1}^n y_i} e^{n(\overline{y}-\lambda_0)}.\end{aligned}$$

It follows that

$$\begin{aligned}-2\log\{\lambda(\boldsymbol{y})\} &= -2\left\{n\overline{y}\log\left(\frac{\lambda_0}{\overline{y}}\right) + n(\overline{y}-\lambda_0)\right\} \\ &= 2n\left\{\overline{y}\log\left(\frac{\overline{y}}{\lambda_0}\right) + \lambda_0 - \overline{y}\right\}.\end{aligned}$$

Here, $p = 1$ and $p_0 = 0$, and so $\nu = 1$. Therefore, by Wilks' theorem, when H_0 is true and n is large,

$$2n\left\{\overline{Y}\log\left(\frac{\overline{Y}}{\lambda_0}\right) + \lambda_0 - \overline{Y}\right\} \sim \chi^2_1.$$

Hence, for a test with approximate significance level α, we reject H_0 if and only if

$$2n\left\{\overline{y}\log\left(\frac{\overline{y}}{\lambda_0}\right) + \lambda_0 - \overline{y}\right\} \geq \chi^2_{1,\alpha}.$$

□

Example 2.39. Suppose that Y_i, $i = 1, \ldots, n$, are iid random variables with the probability mass function given by

$$P(Y = y) = \begin{cases} \vartheta_j, & \text{if } y = j, \quad j = 1, 2, 3; \\ 0, & \text{otherwise,} \end{cases}$$

where ϑ_j are unknown parameters such that $\vartheta_1 + \vartheta_2 + \vartheta_3 = 1$ and $\vartheta_j \geq 0$. Consider testing

$$\begin{aligned} & H_0 : \vartheta_1 = \vartheta_2 = \vartheta_3 \\ \text{against} \quad & H_1 : H_0 \text{ is not true} \end{aligned}$$

We will use the Wilks' theorem to derive the critical region for testing this hypothesis at an approximate significance level α.

Here the full parameter space Θ is two-dimensional because there are only two free parameters, i.e., $\vartheta_3 = 1 - \vartheta_1 - \vartheta_2$ and

$$\Theta = \{\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \vartheta_3)^{\mathrm{T}} : \vartheta_3 = 1 - \vartheta_1 - \vartheta_2, \vartheta_j \geq 0\}.$$

Hence $p = 2$.

The restricted parameter space is zero-dimensional, because under the null hypothesis all the parameters are equal and as they sum up to 1, they all must be equal to $1/3$. Hence

$$\Theta^{\star} = \left\{\boldsymbol{\vartheta} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^{\mathrm{T}}\right\}$$

and so $p_0 = 0$ (zero unknown parameters). That is the number of degrees of freedom of the χ^2 distribution is $\nu = p - p_0 = 2$.

To calculate $\lambda(\boldsymbol{Y})$ we need to find the MLE($\boldsymbol{\vartheta}$) in Θ and in the restricted space $\Theta^{\star}$.

MLE($\boldsymbol{\vartheta}$) in Θ:
The likelihood function is

$$L(\boldsymbol{\vartheta};\boldsymbol{y}) = \vartheta_1^{n_1}\vartheta_2^{n_2}(1-\vartheta_1-\vartheta_2)^{n_3},$$

where n_j is the number of responses equal to j, $j = 1,2,3$. Then, the log-likelihood is

$$l(\boldsymbol{\vartheta};\boldsymbol{y}) = n_1\log(\vartheta_1) + n_2\log(\vartheta_2) + n_3\log(1-\vartheta_1-\vartheta_2).$$

For $j = 1,2$ we have,

$$\frac{\partial l}{\partial\vartheta_j} = \frac{n_j}{\vartheta_j} + \frac{n_3}{1-\vartheta_1-\vartheta_2}(-1).$$

When compared to zero, this gives

$$\frac{\widehat{\vartheta}_j}{n_j} = \frac{\widehat{\vartheta}_3}{n_3} = \gamma.$$

Now, since $\widehat{\vartheta}_1+\widehat{\vartheta}_2+\widehat{\vartheta}_3 = 1$ we obtain $\gamma = \frac{1}{n}$, where $n = n_1+n_2+n_3$. Then the estimates of the parameters are $\widehat{\vartheta}_j = n_j/n$, $j = 1,2,3$.

The second derivatives are

$$\frac{\partial^2 l}{\partial\vartheta_j^2} = -\frac{n_j}{\vartheta_j^2} - \frac{n_3}{\vartheta_3^2}$$
$$\frac{\partial^2 l}{\partial\vartheta_j\vartheta_i} = -\frac{n_3}{\vartheta_3^2}$$

The determinant of the second derivatives is positive for all $\boldsymbol{\vartheta}$ and the element $\frac{\partial^2 l}{\partial\vartheta_1^2}$ is negative for all $\boldsymbol{\vartheta}$, so also for $\widehat{\boldsymbol{\vartheta}}$. Hence, there is a maximum at $\widehat{\boldsymbol{\vartheta}}$ and

$$\mathrm{MLE}(\vartheta_j) = \widehat{\vartheta}_j = \frac{n_j}{n}.$$

MLE($\boldsymbol{\vartheta}$) in $\Theta^\star$:

$$L(\boldsymbol{\vartheta};\boldsymbol{y}) = \left(\frac{1}{3}\right)^{n_1}\left(\frac{1}{3}\right)^{n_2}\left(\frac{1}{3}\right)^{n_3}.$$

That is $\widehat{\boldsymbol{\vartheta}}_0 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^{\mathrm{T}}$. Now, we can calculate $\lambda(\boldsymbol{y})$:

$$\lambda(\boldsymbol{y}) = \frac{L(\widehat{\boldsymbol{\vartheta}}_0;\boldsymbol{y})}{L(\widehat{\boldsymbol{\vartheta}};\boldsymbol{y})} = \frac{\left(\frac{1}{3}\right)^{n_1}\left(\frac{1}{3}\right)^{n_2}\left(\frac{1}{3}\right)^{n_3}}{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2}\left(\frac{n_3}{n}\right)^{n_3}} = \left(\frac{n}{3n_1}\right)^{n_1}\left(\frac{n}{3n_2}\right)^{n_2}\left(\frac{n}{3n_3}\right)^{n_3}.$$

That is

$$-2\log\{\lambda(\boldsymbol{y})\} = -2\sum_{j=1}^{3} n_j \log\left(\frac{n}{3n_j}\right).$$

We reject H_0 at an approximate significance level α if the observed sample belongs to the critical region $\mathcal{R}$, where

$$\mathcal{R} = \left\{\boldsymbol{y}:\ 2\sum_{j=1}^{3} n_j \log\left(\frac{3n_j}{n}\right) \geq \chi^2_{2;\alpha}\right\}.$$

□

2.5.5 Contingency tables

We now show that the usual test for association in contingency tables is a generalized likelihood ratio test and that its asymptotic distribution is an example of Wilks' theorem.

Suppose that we have an $r \times c$ contingency table in which there are Y_{ij} individuals classified in row i and column j, when N individuals are classified independently.

Let ϑ_{ij} be the corresponding probability that an individual is classified in row i and column j, so that $\vartheta_{ij} \geq 0$ and $\sum_{i=1}^{r}\sum_{j=1}^{c}\vartheta_{ij} = 1$.

Then the variables Y_{ij} have a *multinomial* distribution with parameters N and ϑ_{ij} for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$.

Example 2.40. In an experiment 150 patients were allocated to three groups of 45, 45 and 60 patients each. Two groups were given a new drug at different dose levels and the third group received placebo. The responses are as follows:

	Improved	No difference	Worse	$r_i = \sum_{j=1}^{c} y_{ij}$
Placebo	16	20	9	45
Half dose	17	18	10	45
Full dose	26	20	14	60
$c_j = \sum_{i=1}^{r} y_{ij}$	59	58	33	$N = 150$

We are interested in testing the hypothesis that the response to the drug does not depend on the dose level. □

The null hypothesis is that the row and column classifications are independent, and the alternative is that they are dependent.

More precisely, the null hypothesis is $H_0 : \vartheta_{ij} = a_i b_j$ for some $a_i > 0$ and $b_j > 0$, with $\sum_{i=1}^{r} a_i = 1$ and $\sum_{j=1}^{c} b_j = 1$, and the alternative is $H_1 : \vartheta_{ij} \neq a_i b_j$ for at least one pair (i, j) .

The usual test statistic is given by

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(Y_{ij} - E_{ij})^2}{E_{ij}},$$

where

$$E_{ij} = \frac{R_i C_j}{N},$$

$R_i = \sum_{j=1}^{c} Y_{ij}$ and $C_j = \sum_{i=1}^{r} Y_{ij}$.

For large N, $X^2 \sim \chi^2_{(r-1)(c-1)}$ approximately when H_0 is true, and so we reject H_0 at the level of significance α if $X^2 \geq \chi^2_{(r-1)(c-1),\alpha}$.

Now consider the generalized likelihood ratio test. Then the likelihood is

$$L(\boldsymbol{\vartheta}|\boldsymbol{y}) = \frac{N!}{\prod_{i=1}^{r}\prod_{j=1}^{c} y_{ij}!}\prod_{i=1}^{r}\prod_{j=1}^{c}\vartheta_{ij}^{y_{ij}} = A\prod_{i=1}^{r}\prod_{j=1}^{c}\vartheta_{ij}^{y_{ij}},$$

where the coefficient A is the number of ways that N subjects can be divided in rc groups with y_{ij} in the ij-th group.

The log-likelihood is

$$\ell(\boldsymbol{\vartheta};\boldsymbol{y}) = \log(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log(\vartheta_{ij}).$$

We have to maximize this, subject to the constraint $\sum_{i=1}^{r}\sum_{j=1}^{c}\vartheta_{ij} = 1$.

Let Σ' represent the sum over all pairs (i,j) except (r,c). Then we may write

$$\ell(\boldsymbol{\vartheta}|\boldsymbol{y}) = \log(A) + \Sigma' y_{ij}\log(\vartheta_{ij}) + y_{rc}\log\left(1 - \Sigma'\vartheta_{ij}\right).$$

Thus, for $(i,j) \neq (r,c)$, solving the equation

$$\frac{\partial \ell}{\partial \vartheta_{ij}} = \frac{y_{ij}}{\vartheta_{ij}} - \frac{y_{rc}}{1-\Sigma'\vartheta_{ij}} = \frac{y_{ij}}{\vartheta_{ij}} - \frac{y_{rc}}{\vartheta_{rc}} = 0$$

yields $\hat{\vartheta}_{ij}/y_{ij} = \hat{\vartheta}_{rc}/y_{rc} = \gamma$, say.

Since $\sum_{i=1}^{r}\sum_{j=1}^{c}\widehat{\vartheta}_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\gamma = 1$, we have $\gamma = 1/N$, so that the maximum likelihood estimate of ϑ_{ij} is $\hat{\vartheta}_{ij} = y_{ij}/N$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$.

It follows that

$$\ell(\widehat{\boldsymbol{\vartheta}}|\boldsymbol{y}) = \log(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log\left(\frac{y_{ij}}{N}\right).$$

Now, under H_0, we have $\vartheta_{ij} = a_i b_j$ and so

$$\begin{aligned} \ell(\boldsymbol{\vartheta}|\boldsymbol{y}) &= \log(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\{\log(a_i) + \log(b_j)\} \\ &= \log(A) + \sum_{i=1}^{r} r_i \log(a_i) + \sum_{j=1}^{c} c_j \log(b_j). \end{aligned}$$

Now, we maximize this subject to the constraints $\sum_{i=1}^{r} a_i = 1$ and $\sum_{j=1}^{c} b_j = 1$. That is, we maximize

$$\begin{aligned} \ell(\boldsymbol{\vartheta}|\boldsymbol{y}) &= \log(A) + \Sigma' r_i \log(a_i) + r_r \log(a_r) + \Sigma' c_j \log(b_j) + c_c \log(b_c) \\ &= \log(A) + \Sigma' r_i \log(a_i) + r_r \log(1 - \Sigma' a_i) + \Sigma' c_j \log(b_j) + c_c \log(1 - \Sigma' b_j). \end{aligned}$$

Then,

$$\frac{\partial \ell}{\partial a_i} = \frac{r_i}{a_i} - \frac{r_r}{a_r}$$

which, when compared to zero gives

$$\frac{r_i}{\widehat{a}_i} = \frac{r_r}{\widehat{a}_r}$$

or

$$\frac{\widehat{a}_i}{r_i} = \frac{\widehat{a}_r}{r_r} = \gamma.$$

However,

$$1 = \sum_{i=1}^{r} \widehat{a}_i = \sum_{i=1}^{r} r_i \gamma = N\gamma.$$

Hence, $\gamma = 1/N$ and so

$$\widehat{a}_i = \frac{r_i}{N}.$$

Similarly, we obtain the ML estimates for b_j as $\widehat{b}_j = c_j/N$.

Thus, the restricted maximum likelihood estimate of ϑ_{ij} under H_0 is

$$\widehat{\vartheta}_{ij}^{0} = \widehat{a}_i \widehat{b}_j = \frac{r_i}{N}\frac{C_j}{N} = e_{ij}/N$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$.

It follows that

$$\ell(\widehat{\boldsymbol{\vartheta}}_0|\boldsymbol{y}) = \log(A) + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij} \log\left(\frac{e_{ij}}{N}\right).$$

Hence, we obtain

$$-2\log\{\lambda(\boldsymbol{y})\} = -2\log\left\{\frac{L(\widehat{\boldsymbol{\vartheta}}_0|\boldsymbol{y})}{L(\widehat{\boldsymbol{\vartheta}}|\boldsymbol{y})}\right\} = -2\{\ell(\widehat{\boldsymbol{\vartheta}}_0|\boldsymbol{y}) - \ell(\widehat{\boldsymbol{\vartheta}}|\boldsymbol{y})\} = 2\sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log\left(\frac{y_{ij}}{e_{ij}}\right).$$

Here, $p = rc-1$ and $p_0 = (r-1)+(c-1) = r+c-2$, and so $\nu = (r-1)(c-1)$.

Therefore, by Wilks' theorem, when H_0 is true and N is large,

$$2\sum_{i=1}^{r}\sum_{j=1}^{c} Y_{ij}\log\left(\frac{Y_{ij}}{E_{ij}}\right) \underset{approx.}{\sim} \chi^2_{(r-1)(c-1)}.$$

The test statistics X^2 and $-2\log\{\lambda(\boldsymbol{y})\}$ are asymptotically equivalent, that is, they differ by quantities that tend to zero as $N \to \infty$.

To see this, first note that $y_{ij} - e_{ij}$ is small relative to y_{ij} and e_{ij} when N is large.

Thus, since $y_{ij} = e_{ij} + (y_{ij} - e_{ij})$, we may write (by Taylor series expansion of $\log(1+x)$ around x, cut after second order term)

$$\log\left(\frac{y_{ij}}{e_{ij}}\right) = \log\left(1 + \frac{y_{ij}-e_{ij}}{e_{ij}}\right) \simeq \frac{(y_{ij}-e_{ij})}{e_{ij}} - \frac{1}{2}\frac{(y_{ij}-e_{ij})^2}{e_{ij}^2}.$$

Hence, we have

$$\begin{aligned} y_{ij}\log\left(\frac{y_{ij}}{e_{ij}}\right) &\simeq \{e_{ij} + (y_{ij}-e_{ij})\}\left\{\frac{(y_{ij}-e_{ij})}{e_{ij}} - \frac{1}{2}\frac{(y_{ij}-e_{ij})^2}{e_{ij}^2}\right\} \\ &\simeq (y_{ij}-e_{ij}) + \frac{1}{2}\frac{(y_{ij}-e_{ij})^2}{e_{ij}}. \end{aligned}$$

Since $\sum_{i=1}^{r}\sum_{j=1}^{c}(y_{ij}-e_{ij}) = 0$, it follows that

$$-2\log\{\lambda(\boldsymbol{y})\} \simeq \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(y_{ij}-e_{ij})^2}{e_{ij}}.$$

So we now see why we use the X^2 test statistic.

Example 2.41. continued.
Now we will test the hypothesis from the previous example. The table of e_{ij} values is following

	Improved	No difference	Worse
Placebo	17.7	17.4	9.9
Half dose	17.7	17.4	9.9
Full dose	23.6	23.2	13.2

This gives

$$X^2_{obs.} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(y_{ij}-e_{ij})^2}{e_{ij}} = 1.417.$$

The critical value is $\chi^2_{4;0.05} = 9.488$, hence there is no evidence to reject the null hypothesis saying that the different responses are independent of the levels of the new drug and placebo.

We obtain the same conclusion using the critical region derived from the Wilks' theorem, which is

$$\mathcal{R} = \left\{\boldsymbol{y} : 2\sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log\left(\frac{y_{ij}}{e_{ij}}\right) \geq \chi^2_{\nu;\alpha}\right\}.$$

Here

$$2\sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log\left(\frac{y_{ij}}{e_{ij}}\right) = 1.42.$$