(5.5) Wilks' Theorem

Theorem 5.2 (no proof) For testing $H_0: \underset{\sim}{\theta} \in \omega$

vs. $H_1: \underset{\sim}{\theta} \in \Omega \setminus \omega$

where $\dim \Omega = p$ and where $\dim \omega = p_0$

ie p ~~independent~~ parameters and $p - p_0$ is the number of (independent) constraints imposed by H_0

$$-2\log \Lambda(\underset{\sim}{Y}) = 2\left(\ell(\underset{\sim}{\hat\theta};\underset{\sim}{Y}) - \ell(\underset{\sim}{\hat\theta}_0;\underset{\sim}{Y})\right) \sim \chi^2_{p-p_0}$$

twice difference in maximised value of the loglikelihood

For large n, under H_0 and 'regularity conditions'

Ex 1(i) revisited ~~$p=1$~~ $(p=1)$ $H_0: \lambda = \lambda_0$ so $p_0 = 0$ $(p - p_0 = 1)$

$$\Lambda(\underset{\sim}{y}) = \left(\lambda_0/\hat\lambda\right)^{na} e^{(\hat\lambda - \lambda_0)\sum y_i} \qquad \hat\lambda = a/\bar{y}$$

so

$$-2\log\Lambda(\underset{\sim}{Y}) = -2\left[na\log\lambda_0 - na\log\hat\lambda\right] - 2(\hat\lambda - \lambda_0)n\bar{Y}$$

opposite signs

$$= 2na\left[\log\left(\frac{a/\lambda_0}{\bar{Y}}\right) + \left(\frac{\bar{Y}}{(a/\lambda_0)} - 1\right)\right] \sim \chi^2_1$$

in terms of the ratio of sample mean $\bar{Y}$ to hypothesised mean $a/\lambda_0 = \mu_0$ (under H_0).

reject H_0 if

$$2na\left(\log(\mu_0/\bar{y}) + (\bar{y}/\mu_0 - 1)\right) > \chi^2_{1;\alpha}$$

For a test of size α.

Simple & no problems with asymmetry but must have large enough n.

Similarly more complicated examples

Ex 1(ii) $p=2$ $\quad H_0: \lambda = \lambda_0$ as before
but now $\dim \omega = p_0 = 1$
($p - p_0 = 1$ as before)
independent restrictions

$$L(\lambda, a; \underline{y}) = \lambda^{na} e^{-\lambda \sum y_i} \left(\prod y_i\right)^{a-1} \Big/ [\Gamma(a)]^n$$

so

$$\ell(\lambda, a; \underline{y}) = na \log \lambda - \lambda \sum y_i + (a-1) \sum \log y_i - n \log \Gamma(a)$$

$$-2 \log \Lambda(\underline{y}) = 2\left[\ell(\hat{\lambda}, \hat{a}; \underline{y}) - \ell(\lambda_0, \hat{a}_0; \underline{y})\right] \qquad \hat{\lambda} = \hat{a}/\bar{y}$$

($\hat{a}, \hat{a}_0$ obtained numerically)

$$= 2\left[n\hat{a} \log \hat{\lambda} - n\hat{a}_0 \log \lambda_0 - n\bar{y}(\hat{\lambda} - \lambda_0) + (\hat{a} - \hat{a}_0) \sum \log y_i - n \log \frac{\Gamma(\hat{a})}{\Gamma(\hat{a}_0)}\right]$$

$\sim \chi^2_1$ under H_0

NB if $\hat{a} = \hat{a}_0$, then as in Ex 1(i) (with $\hat{a}$ subs. for a)

so again reject H_0 if this statistic exceeds $\chi^2_{1;\alpha}$.

Ex 2 revisited ($p=3$) $\quad H_0: \mu_1 = \mu_2$ $\quad$ ω has dimension $p_0 = 2$ or $p - p_0 = 1$ independent restrictions

$-2 \log \Lambda(\underline{y}) \sim \chi^2_1$ for asymptotic test

but exact test using $T \sim t_{n-2}$

Similarly Exs 3 & 4

(5.6) Goodness-of-Fit

(Webnotes: Contingency Tables)
(independence in)

Starting point:

data $y_1, y_2, \ldots, y_n$

observed values of $Y_1, Y_2, \ldots, Y_n$

which are independent with identical distribution (IID)

$f_Y(y\,;\,\underset{\sim}{\theta})$ (PMF or pdf) ~~with say $\underset{\sim}{\theta}$ has q parameters~~.

ie a random sample from a distribution of known form but with say q unknown parameters.

Can we test this assumption?

test H_0: ~~[illegible]~~ random sample from $f_Y(y\,;\,\underset{\sim}{\theta})$

Now $\dim \omega = q$ but what is Ω?

ie what is the alternative hypothesis?

the set of unknown distributions have an infinite number of parameters but we know we can get a perfect fit to our data with just n parameters.

$\therefore$ Summarise data by grouping into K classes or cells $C_1, C_2, \ldots, C_K$ ~~(usually spanning R or R+)~~ (usually intervals spanning the range of Y: the support of $f_Y(y\,;\,\underset{\sim}{\theta})$)

to give observed frequencies n_k with $\sum n_k = n$.

$\underset{\sim}{n} = (n_1, \ldots, n_K)^T \qquad k=1,\ldots,K$

compare with expected frequencies under H_0

~~[illegible]~~

What is the likelihood of these data?

Let $\pi_k = P[Y \in C_k] \quad k=1,\ldots,K \quad \sum \pi_k = 1$

then $\underset{\sim}{\pi} = (\pi_1, \ldots, \pi_K)^T$

the joint distribution of the r.v's (whose observed values are the observed frequencies $\underset{\sim}{n}$) is multinomial

(still assuming independence)

$$L(\underset{\sim}{\pi}\,;\,\underset{\sim}{n}) = \prod_{k=1}^{K} \frac{\pi_k^{n_k}}{n_k!} \cdot n!$$

(binomial $(n;p)$ is special case with $K=2$)

$\pi_1 = p, \; \pi_2 = 1-p$

We follow LR procedures and estimate $\underset{\sim}{\pi}$ by maximum likelihood

(i) under H_0 ie in ω with dimension q

(ii) over $\Omega = \{\underset{\sim}{\pi} : \sum \pi_k = 1,\ \pi_k \geq 0\ k=1,\ldots,K\}$

$\underline{\dim \Omega = K-1}$ (a '<u>simplex</u>')

<u>(i) under H_0</u> $\quad \pi_k = \pi_k(\underset{\sim}{\theta})$ (determined by integrating or summing y over C_k)

so the restricted MLEs are

$$\hat{\pi}_{k0} = \pi_k(\hat{\underset{\sim}{\theta}}) \qquad \underline{\text{estimated probabilities}}$$

where $\hat{\underset{\sim}{\theta}}$ is the MLE of $\underset{\sim}{\theta}$ (estimated from the original data?)

and hence $\hat{n}_k = n\pi_k(\hat{\underset{\sim}{\theta}})$ $\quad k=1,\ldots,K$

(or E_k) are the <u>expected frequencies</u>

<u>(ii) over Ω</u>; ~~...~~ we maximise $L(\underset{\sim}{\pi}; \underset{\sim}{n})$ ~~subject to $\sum \pi_k = 1$~~.

~~Lagrange multiplier λ: maximise L ... 1~~

or $\ell(\underset{\sim}{\pi}; \underset{\sim}{n}) = \sum n_k \log \pi_k \quad \underbrace{+ \log n! - \sum \log n_k!}_{\text{const.}}$

subject to $\sum \pi_k = 1$

∴ introduce a Lagrange multiplier λ & maximise $\ell^* = \ell + \lambda(1 - \sum \pi_k)$

(without constraints)

$$\frac{\partial \ell^*}{\partial \pi_k} = \frac{n_k}{\pi_k} - \lambda = 0 \quad (1) \quad k=1,\ldots,K$$

$$\frac{\partial \ell^*}{\partial \lambda} = 1 - \sum \pi_k = 0 \quad \text{(giving the constraint back again)}$$

The only solution is $\hat{\pi}_k = n_k/n \quad k=1,\ldots,K$

(as (1) $\Rightarrow n_k/\pi_k$ is constant over k)

and $n\hat{\pi}_k = n_k$ ~~...~~ (or O_k) are the <u>observed frequencies</u>

$$\Lambda(\underset{\sim}{y}) = \frac{L(\hat{\underset{\sim}{\pi}}_0; \underset{\sim}{n})}{L(\hat{\underset{\sim}{\pi}}; \underset{\sim}{n})} \quad \text{or the Wilks' statistic}$$

$$-2\log \Lambda(\underset{\sim}{y}) = 2\left[\ell(\hat{\underset{\sim}{\pi}}; \underset{\sim}{n}) - \ell(\hat{\underset{\sim}{\pi}}_0; \underset{\sim}{n})\right)$$

going as $\hat{\pi}_k = n_k/n$ $\quad \hat{\pi}_{k0} = \pi_k(\hat{\theta})$

$$= 2\Big[\sum n_k \log n_k - \underbrace{\sum n_k \log n}_{\cancel{n \log n}} - \underbrace{\sum n_k \log \pi_k(\hat{\theta})}_{-\sum n_k \log \hat{n}_k + \underbrace{\sum n_k \log n}_{\cancel{n \log n}}}\Big]$$

$$= 2\sum_{k=1}^{K} n_k \log(n_k/\hat{n}_k)$$

not the usual 'Pearson chi square'

$$X^2 = \sum_{k=1}^{K} \frac{(n_k - \hat{n}_k)^2}{\hat{n}_k}$$

In this example we call Wilks' statistic Y^2 (to contrast with X^2)

However if we write

$$n_k = \hat{n}_k + (n_k - \hat{n}_k)$$

and note that $(n_k - \hat{n}_k)$ is small relative to n_k and $\hat{n}_k$ when n is large we can ~~e~~ approximate using the first 2 terms in an expansion

$$\log(n_k/\hat{n}_k) = \log\left(1 + \frac{n_k - \hat{n}_k}{\hat{n}_k}\right)$$

$$\doteq \frac{n_k - \hat{n}_k}{\hat{n}_k} - \frac{1}{2}\left(\frac{n_k - \hat{n}_k}{\hat{n}_k}\right)^2$$

ie

$$n_k \log(n_k/\hat{n}_k) \doteq [\hat{n}_k + (n_k - \hat{n}_k)] \log\left[1 + \frac{n_k - \hat{n}_k}{\hat{n}_k}\right]$$

$$\doteq (n_k - \hat{n}_k) + \frac{1}{2}\frac{(n_k - \hat{n}_k)^2}{n_k}$$

and when we sum over k we get

$$Y^2 = 2\sum_k n_k \log(n_k/\hat{n}_k) \doteq \underbrace{2\sum_k (n_k - \hat{n}_k)}_{0} + \underbrace{\sum_k \frac{(n_k - \hat{n}_k)^2}{n_k}}_{X^2}$$

d.f = no. of classes − no. of ~~estimated~~ parameters estimated by MLE − 1

so ~~the~~ X^2 is an approximation to ~~the Wilks~~ Y^2 (asymptotically equivalent ~~statistic~~)

~~We reject Ho~~

IF n large, Wilks' theorem gives

$$Y^2 (\text{or } X^2) \sim \chi^2_{p-p_0}$$

but $p = K-1$, $p_0 = q$ so this is χ^2 as in Year 1