Hypothesis Tests and Confidence Regions Using the Likelihood

F. P. Treasure

13-May-05
3020a

Maximum Likelihood:

The likelihood function for a model parameterised by θ (a vector in a p-dimensional space Θ) given observed data vector x is $L(\theta|x)$ which we abbreviate to $L(\theta)$. The log-likelihood $S(\theta)$ is defined by $S(\theta) := \log L(\theta)$.

The value of θ which maximizes $S(\theta)$ is $\hat{\theta}$ and is the *maximum likelihood estimate* of θ :

$$S(\hat{\theta}) = \max_{\theta \in \Theta} S(\theta) \ .$$

$\hat{\theta}$ is normally found by solving the score equations $S'(\hat{\theta}) = 0$, where a prime denotes differentiation.

If θ is constrained to a q-dimensional subspace Θ_0 of Θ then the value of θ maximizing $S(\theta)$ in that subspace is $\tilde{\theta}$:

$$S(\tilde{\theta}) = \max_{\theta \in \Theta_0} S(\theta)$$

and $\tilde{\theta}$ can generally be found using Lagrange multipliers.

Hypothesis tests

The (non-negative) difference $S(\hat{\theta}) - S(\tilde{\theta})$ is the reduction in the log-likelihood due to constraining the space from Θ to Θ_0 . If the data does not support $\theta \in \Theta_0$ then we would expect $S(\hat{\theta}) - S(\tilde{\theta})$ to be relatively large and vice-versa.

Wilks's lemma tells us that:

$$2\left[S(\hat{\theta}) - S(\tilde{\theta})\right] \sim \text{chisquare}\,(p-q)$$

(provided $\theta \in \Theta_0$) and so we can compare $2\left[S(\hat{\theta}) - S(\tilde{\theta})\right]$ with the chisquare $(p-q)$ distribution and obtain a size α hypothesis test for the null hypothesis $\theta \in \Theta_0$:

$$\text{accept hypothesis } \theta \in \Theta_o \text{ if } S(\hat{\theta}) - S(\tilde{\theta}) \leqslant \frac{1}{2} C_{p-q,1-\alpha} \tag{1}$$

where $C_{m,\gamma}$ is the γth quantile of a chisquare(m) distribution.

Confidence Regions and Intervals

A p-dimensional $1-\alpha$ confidence *region* can be constructed by letting Θ_0 consist of the single point θ_0 and including θ_0 in the confidence region if the hypothesis test $\theta \in \Theta_0$ – equivalently: $\theta = \theta_0$ – is not rejected by rule (1). The confidence region is therefore (noting here that $q = 0$ and $\tilde{\theta} = \theta_0$):

$$\left\{ \theta_0 : S(\hat{\theta}) - S(\theta_0) \leqslant \frac{1}{2} C_{p-q,1-\alpha} \right\} .$$

A confidence *interval* is a one-dimensional confidence region and is obtained when either θ is one-dimensional ($p = 1$) or we are interested in a single component of θ . In the latter case we partition θ as $\theta = [\beta\ \psi]^{\mathsf{T}}$ where β is a scalar (parameter of interest) and ψ is $(p-1)$-dimensional (the nuisance parameters). The maximum likelihood estimate $\hat{\theta}$ is now $[\hat{\beta}\ \hat{\psi}]^{\mathsf{T}}$. The symbols ψ and $\hat{\psi}$ can be ignored if θ is one-dimensional.

The confidence interval will be of form $L \leqslant \beta_0 \leqslant U$ and is given by:

$$\left\{ \beta_0 : S(\hat{\beta}, \hat{\psi}) - S(\beta_0, \tilde{\psi}) \leqslant \frac{1}{2} C_{1,1-\alpha} \right\}$$

where $\tilde{\psi}$ is defined by $S(\beta_0, \tilde{\psi}) = \max_{[\beta\ \psi]^{\mathsf{T}} \in \Theta, \beta = \beta_0} S(\beta, \psi)$.

Other Tests Based on the Likelihood

The above tests and confidence regions are based on the difference in log-likelihoods and are referred to as *likelihood-ratio* tests etc. They are (in my view) the best ones to use. For historical and computational reasons two other approaches are commonly seen. They are based on approximating the log-likelihood function by a quadratic and are both asymptotically equivalent to likelihood-ratio methods. The tests are in practice only as good as the quadratic approximation (usually good enough)

I shall present the approximate methods using the simplest case: a hypothesis test for a single scalar parameter (that is: $p = 1$ and $q = 0$). In theoretical work the expectation $\mathcal{E}S''(\theta)$ is often used instead of the observed $S''(\theta)$: this is rarely practicable (and arguably not desirable) in survival analysis.

The ***Wald*** Test

The log-likelihood is approximated by a quadratic at $\beta = \hat{\beta}$. The statistic for testing the null hypothesis that $\beta = \beta_0$ is

$$-\left(\hat{\beta} - \beta_0\right)^2 S''(\hat{\beta})$$

which is compared with the chisquare(1) distribution. Many computer programs report the reciprocal of the square root of $S''(\hat{\beta})$ as the estimated standard deviation of $\hat{\beta}$ (the 'standard error').

The *Score* Test

The log-likelihood is approximated by a quadratic at $\beta = \beta_0$. This has the huge computational advantage that the log-likelihood does not have to be maximized. The test statistic is:

$$\frac{\left[S'\left(\beta_0\right)\right]^2}{-S''\left(\beta_0\right)}\ , \tag{2}$$

again, compared with the chisquare(1) distribution.

Exercise (hard(ish)): show that the score test applied to a proportional hazards model of a two group comparison gives the log-rank test. Hint: ignore the denominator in both (2) and the log-rank statistic as they merely normalise the variance to unity – concentrate on showing the numerators are proportional.

References

1. Therneau T. M. and Grambsch P. M. (2000) *Modelling Survival Data – Extending the Cox Model*. Springer-Verlag [see chapter three for useful summary and examples and an illustration of what to do when the quadratic approximation breaks down]

2. Wilks S. S. (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* 9:60-62