Testing Null Hypotheses[1]

STA431 Spring

[1] See last slide for copyright information.

Vector of MLEs is Asymptotically Normal

That is, Multivariate Normal

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \stackrel{d}{\rightarrow} \mathbf{t} \sim N_k\left(\mathbf{0}, \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}\right)$$

Approximating the asymptotic covariance matrix $\frac{1}{n}\boldsymbol{\mathcal{I}}(\boldsymbol{\theta})^{-1}$ with $\widehat{\mathbf{V}}_n = \mathbf{H}^{-1}$ yields confidence intervals for the parameters, and

- Z-tests
- Wald tests.
- Indirectly, Likelihood Ratio tests.

Z-tests

Have $Z_j = \frac{\widehat{\theta}_j - \theta_j}{se_{\widehat{\theta}_j}}$ approximately standard normal, where $se_{\widehat{\theta}_j}$ is the square root of the jth diagonal element of $\widehat{\mathbf{V}}_n$.

Test $H_0 : \theta_j = \theta_0$ using

$$Z = \frac{\widehat{\theta}_j - \theta_0}{se_{\widehat{\theta}_j}}$$

And Wald Tests for $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$

Based on $(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2(p)$

$$W_n = (\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{h})^\top \left(\mathbf{L}\widehat{\mathbf{V}}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{h})$$

$\widehat{\boldsymbol{\theta}}_n \stackrel{.}{\sim} N_p(\boldsymbol{\theta}, \mathbf{V_n})$ so if H_0 is true, $\mathbf{L}\widehat{\boldsymbol{\theta}}_n \stackrel{.}{\sim} N_r(\mathbf{h}, \mathbf{L}\mathbf{V}_n\mathbf{L}^\top)$.

Thus $(\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{h})^\top \left(\mathbf{L}\mathbf{V}_n\mathbf{L}^\top\right)^{-1} (\mathbf{L}\widehat{\boldsymbol{\theta}}_n - \mathbf{h}) \stackrel{.}{\sim} \chi^2(r)$.

Substitute $\widehat{\mathbf{V}}_n$ for $\mathbf{V}_n$.

The Wtest function

`source("https://www.utstat.toronto.edu/brunner/openSEM/fun/Wtest.txt")`

```
Wtest = function(L,Tn,Vn,h=0) # H0: L theta = h
# Tn is estimated theta, usually a vector.
# Vn is the estimated asymptotic covariance matrix of Tn.
# For Wald tests based on numerical MLEs, Tn = theta-hat,
# and Vn is the inverse of the Hessian of the minus log
# likelihood.
    {
    Wtest = numeric(3)
    names(Wtest) = c("W","df","p-value")
    r = dim(L)[1]
    W = t(L%*%Tn-h) %*% solve(L%*%Vn%*%t(L)) %*%
         (L%*%Tn-h)
    W = as.numeric(W)
    pval = 1-pchisq(W,r)
    Wtest[1] = W; Wtest[2] = r; Wtest[3] = pval
    Wtest
    } # End function Wtest
```

Likelihood Ratio Tests

$$x_1, \ldots, x_n \overset{i.i.d.}{\sim} F_\theta,\ \theta \in \Theta,$$
$$H_0 : \theta \in \Theta_0 \text{ v.s. } H_A : \theta \in \Theta \cap \Theta_0^c,$$

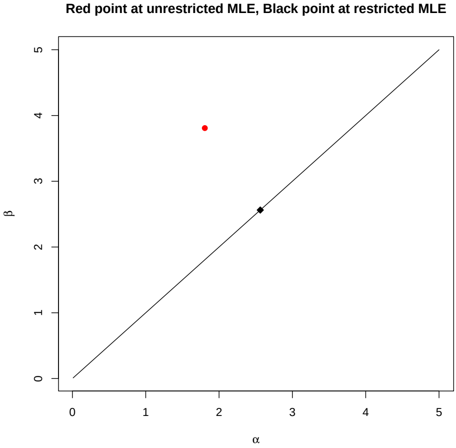
$$\begin{aligned} G^2 &= -2 \ln \left(\frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} \right) \\ &= -2 \ln \left(\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \right) \end{aligned}$$

Under H_0, G^2 has an approximate chi-square distribution for large n. Degrees of freedom = number of (non-redundant, linear) equalities specified by H_0. Reject when G^2 is large.

Example: Testing $H_0 : \alpha = \beta$ for data from a Gamma distribution

$H_0 : \theta \in \Theta_0$ v.s. $H_A : \theta \in \Theta \cap \Theta_0^c$

- $\Theta = \{(\alpha, \beta) : \alpha > 0, \beta > 0\}$
- $\Theta_0 = \{(\alpha, \beta) : \alpha = \beta > 0\}$



Functions

$-\ell(\alpha,\beta) = n\alpha\ln\beta + n\ln\Gamma(\alpha) + \frac{1}{\beta}\sum_{i=1}^{n} x_i - (\alpha-1)\sum_{i=1}^{n}\ln x_i$

```
gmll = function(theta,datta)
     {
       aa =  theta[1]; bb =  theta[2]
       nn = length(datta); sumd = sum(datta)
       sumlogd = sum(log(datta))
       value = nn*aa*log(bb) + nn*lgamma(aa) + sumd/bb - (aa-1)*sumlogd
       return(value)
     } # End function gmll


# gmll0 is minus LL gamma log likelihood with alpha=beta
gmll0 = function(alpha,datta) gmll(c(alpha,alpha),datta)
```

```
> # Unrestricted MLE
> gsearch = optim(par=c(momalpha,mombeta), fn = gmll,
+                 method = "L-BFGS-B", lower = c(0,0), hessian=TRUE, datta=d)
> gsearch

$par
[1] 1.805930 3.808674

$value
[1] 142.0316

$counts
function gradient
       9        9

$convergence
[1] 0

$message
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

$hessian
         [,1]      [,2]
[1,] 36.69402 13.127928
[2,] 13.12793  6.224773
```

Restricted MLE

Restricted by $H_0 : \alpha = \beta$

```
> gsearch0 = optim(par=mean(thetahat), fn = gmll0,
+                  method = "L-BFGS-B", lower = 0,  datta=d)
> gsearch0
$par
[1] 2.562371

$value
[1] 144.1704

$counts
function gradient
       6        6

$convergence
[1] 0

$message
[1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```

Likelihood Ratio Test

$G^2 = -2\ln\left(\frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})}\right) = 2\left(-\ln L(\widehat{\theta}_0) - (-\ln L(\widehat{\theta}))\right)$

```
> Gsq = 2*(gsearch0$value-gsearch$value); Gsq
[1] 4.277603
> pval = 1-pchisq(Gsq,df=1); pval
[1] 0.03861777
```

```
> thetahat
alpha-hat  beta-hat
 1.805930  3.808674
```

Wald test for comparison

Likelihood ratio test yielded $G^2 = 4.278$, $p = 0.0386$

```
> source("https://www.utstat.toronto.edu/brunner/openSEM/fun/Wtest.txt")
> LL = rbind(c(1,-1))
> Wtest(LL,Tn=thetahat,Vn=Vhat_n)
         W         df    p-value
3.25110020 1.00000000 0.07137553
```

Comparing Likelihood Ratio and Wald Tests in General

- Asymptotically equivalent under H_0, meaning $(W_n - G^2_n) \stackrel{p}{\rightarrow} 0$
- Under the alternative hypothesis,
  - Both have the same approximate distribution (non-central chi-square).
  - Both go to infinity as $n \rightarrow \infty$.
  - But values are not necessarily close.
- Likelihood ratio test tends to get closer to the right Type I error probability for small samples.
- Wald can be more convenient when testing lots of hypotheses, because you only need to fit the model once.
- Wald can be more convenient if it's a lot of work to write the restricted likelihood.

Z-test of $H_0 : \beta = 3$

```
> se = sqrt(Vhat_n[2,2])
> # Assignning names because otherwise everything is labelled "betahat"
> z = (thetahat[2]-3)/se; names(z) = "Z statistic"; z
Z statistic
  0.9996297
pval = 2*(1-pnorm(abs(z))); names(pval) = "p-value"; pval
  p-value
0.3174897
```

Copyright Information