

# Large-Sample Likelihood Ratio Tests

We will use the following hypothesis-testing framework. The data are  $Y_1, \dots, Y_n$ . The distribution of these independent and identically distributed random variables depends on the parameter  $\theta$ , and we are testing a null hypothesis  $H_0$  using a large sample likelihood ratio test.

$$\begin{aligned} Y_1, \dots, Y_n &\overset{i.i.d.}{\sim} P_\theta, \theta \in \Theta, \\ H_0 : \theta &\in \Theta_0 \text{ v.s. } H_A : \theta \in \Theta \cap \Theta_0^c, \end{aligned}$$

The data have likelihood function

$$\ell(\theta) = \prod_{i=1}^n f_\theta(y_i),$$

where  $f_\theta(y_i)$  is the joint density or probability mass function evaluated at  $y_i$ .

Let  $\hat{\theta}$  denote the usual Maximum Likelihood Estimate (MLE). That is, it is the parameter value for which the likelihood function is greatest, over all  $\theta \in \Theta$ . And, let  $\hat{\hat{\theta}}$  denote the *restricted* MLE. The restricted MLE is the parameter value for which the likelihood function is greatest, over all  $\theta \in \Theta_0$ . This MLE is *restricted* by the null hypothesis  $H_0 : \theta \in \Theta_0$ . It should be clear that  $\ell(\hat{\hat{\theta}}) \leq \ell(\hat{\theta})$ .

Tests will be based on

$$\begin{aligned} G &= -2 \log \left( \frac{\max_{\theta \in \Theta_0} \prod_{i=1}^n f_\theta(y_i)}{\max_{\theta \in \Theta} \prod_{i=1}^n f_\theta(y_i)} \right) \\ &= -2 \log \left( \frac{\prod_{i=1}^n f_{\hat{\hat{\theta}}}(y_i)}{\prod_{i=1}^n f_{\hat{\theta}}(y_i)} \right) \\ &= -2 \log \frac{\ell(\hat{\hat{\theta}})}{\ell(\hat{\theta})} \\ &= [-2 \log \ell(\hat{\hat{\theta}})] - [-2 \log \ell(\hat{\theta})]. \end{aligned}$$

Thus, the test statistic  $G$  is the *difference* between two  $-2 \log$  likelihood functions. If you are doing the maximum likelihood estimation numerically, this means you can minimize  $-2 \log \ell(\theta)$  twice, first over all  $\theta \in \Theta$ , and then over all  $\theta \in \Theta_0$ . The test statistic is the difference between the two minimum values.

If the null hypothesis is true, then the test statistic  $G$  has, if the sample size is large, an approximate chisquare distribution, with degrees of freedom equal to the difference of the *dimension* of  $\Theta$  and  $\Theta_0$ . For example, if the null hypothesis is that 4 elements of  $\theta$  equal zero, then the degrees of freedom are equal to 4. More generally, if the null hypothesis imposes  $k$  linear restrictions on  $\theta$ , then the degrees of freedom equal  $k$ .

Think of the usual normal multiple regression model. Here,  $\theta = (\boldsymbol{\beta}, \sigma^2)$ . Consider the null hypothesis  $H_0 : \mathbf{L}\boldsymbol{\beta} = \boldsymbol{\gamma}$ , where  $\mathbf{L}$  is a  $k \times p$  matrix. This null hypothesis imposes  $k$  linear restrictions on the parameter, one for each row of  $\mathbf{L}$ . The dimension of  $\Theta$  is  $p + 1$ ; the dimension of  $\Theta_0$  is  $p + 1 - k$ .

The  $p$ -value associated with the test statistic  $G$  is  $Pr\{X > G\}$ , where  $X$  is a chisquare random variable with  $k$  degrees of freedom. If  $p < \alpha$ , we reject  $H_0$  and call the results “statistically significant.”

**Example** Let  $X_1, \dots, X_{n_1}$  be a random sample from a Gamma distribution with parameters  $\alpha_1$  and  $\beta_1$ . Independently of the  $X$  values, let  $Y_1, \dots, Y_{n_2}$  be a random sample from a Gamma distribution with parameters  $\alpha_2$  and  $\beta_2$ . We will test  $H_0 : \alpha_1 = \alpha_2$ .

The parameter for this problem is  $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ , and the null hypothesis imposes one linear restriction on the parameter. So, the degrees of freedom of the large-sample likelihood ratio chisquare test will equal one.

The likelihood function is

$$\ell(\theta) = \prod_{i=1}^{n_1} \frac{1}{\beta_1^{\alpha_1} \Gamma(\alpha_1)} e^{x_i/\beta_1} x_i^{\alpha_1-1} \prod_{j=1}^{n_2} \frac{1}{\beta_2^{\alpha_2} \Gamma(\alpha_2)} e^{y_j/\beta_2} y_j^{\alpha_2-1}.$$

Any time we are finding the unrestricted MLE for this kind of problem, we can do it in two steps. When we are maximizing over  $(\alpha_1, \beta_1)$ , then the material from Sample 2 is a constant and can be ignored. And when we are maximizing over  $(\alpha_2, \beta_2)$ , the material from Sample 1 is a constant. Then,

$$-2 \log \ell(\hat{\theta}) = -2 \log[l_1(\hat{\theta}_1)l_2(\hat{\theta}_2)] = [-2 \log l_1(\hat{\theta}_1)] + [-2 \log l_2(\hat{\theta}_2)].$$

Thus, to obtain -2 time the overall log likelihood evaluated at the MLE, we can do the minimization of -2 times the log likelihood separately for each sample, and then add the minimum values.

For the restricted MLE, we have a 3-parameter problem, and we need to minimize the following expression over  $(\alpha, \beta_1, \beta_2)$ :

$$\begin{aligned} -2 \log \ell(\alpha, \beta_1, \beta_2) &= 2\alpha(n_1 \log \beta_1 + n_2 \log \beta_2) \\ &+ 2(n_1 + n_2) \log \Gamma(\alpha) \\ &+ \frac{2}{\beta_1} \sum_{i=1}^{n_1} x_i + \frac{2}{\beta_2} \sum_{j=1}^{n_2} y_j \\ &- 2(\alpha - 1) \left( \sum_{i=1}^{n_1} \log x_i + \sum_{j=1}^{n_2} \log y_j \right). \end{aligned}$$

Now we'll calculate  $G$  and the  $p$ -value for a data set. The data come in two columns; the first column indicates group membership, and the second column contains the values of the gamma random variables. Here is the R program `gammatest.R`.

```
#####
# gammatest.R  Two independent Gamma samples. Test H0: alpha1=alpha2 #
#              Execute with  source("gammatest.R")                #
#####
#
# Define functions
#
GamLL <- function(theta,xx) # -2 Log likelihood for a single Gamma distribution
                          # Data are in xx
{
  alpha <- theta[1] ; beta <- theta[2]
  n <- length(xx) ; sumx <- sum(xx) ; sumlogx <- sum(log(xx))
  GamLL <- 2*n*alpha*log(beta) + 2*n*lgamma(alpha) + 2*sumx/beta -
          2*(alpha-1)*sumlogx
  # Fix up to avoid negative alpha, beta
  # The nlm function will replace Inf with the largest machine value
  if(alpha <= 0) GamLL <- Inf ; if(beta <= 0) GamLL <- Inf
  GamLL
} # end of function GamLL

alphaeq <- function(theta,xx,yy) # Restricted MLE for two independent Gamma
                                # samples. H0: alpha1=alpha2
{
  alpha <- theta[1] ; beta1 <- theta[2] ; beta2 <- theta[3]
  n1 <- length(xx) ; n2 <- length(yy)
  sumx <- sum(xx) ; sumy <- sum(yy)
  sumlogx <- sum(log(xx)) ; sumlogy <- sum(log(yy))
  alphaeq <- 2*alpha*(n1*log(beta1)+n2*log(beta2)) +
            2*(n1+n2)*lgamma(alpha) +
            (2/beta1)*sumx + (2/beta2)*sumy -
            2*(alpha-1)*(sumlogx+sumlogy)
  # Fix up to avoid negative alpha, betas
  # The nlm function will replace Inf with the largest machine value
  if(alpha <= 0) alphaeq <- Inf
  if(beta1 <= 0) alphaeq <- Inf ; if(beta2 <= 0) alphaeq <- Inf
  alphaeq
} # End of function alphaeq

#
# Read the data
#
datta <- read.table("gammatest.dat")
datta[1:10,] # Look at first 10 rows
datta[120:125,] # Last 6 rows
x <- datta[,2][datta[,1]==1] # x gets numbers in the second column such
                              # that the number in the first column equals one.
y <- datta[,2][datta[,1]==2]
#
```

```

# Compute and print MLEs
#
#   First, unrestricted. Get starting values with method of moments
#
xbar <- mean(x) ; s2 <- var(x)
astart <- xbar^2/s2 ; bstart <- s2/xbar
samp1 <- nlm(GamLL,c(astart,bstart),xx=x)[1:2]
      # First element is minimum value
      # Second element is pair (alphahat,betahat)
xbar <- mean(y) ; s2 <- var(y)
astart <- xbar^2/s2 ; bstart <- s2/xbar
samp2 <- nlm(GamLL,c(astart,bstart),xx=y)[1:2]
alphahat1 <- samp1$estimate[1] ; betahat1 <- samp1$estimate[2]
alphahat2 <- samp2$estimate[1] ; betahat2 <- samp2$estimate[2]
unrest <- samp1$minimum + samp2$minimum # -2LL at unrestricted MLE
#
#   Now restricted. Use crude modification of the unrestricted MLEs
#   as starting values.
#
astart <- (samp1$estimate[1]+samp2$estimate[1])/2 # Mean of unrestricted MLEs
b1start <- samp1$estimate[2] ; b2start <- samp1$estimate[2]
together <- nlm(alphaeq,c(astart,b1start,b2start),xx=x,yy=y)[1:2]
alphahathat <- together$estimate[1] ; betahathat1 <- together$estimate[2]
betahathat2 <- together$estimate[3]
rest <- together$minimum # -2LL at restricted MLE
#
cat("\n")
cat("Unrestricted MLE: \n")
cat("  alphahat1 = ",alphahat1," betahat1 = ",betahat1,"\n")
cat("  alphahat2 = ",alphahat2," betahat2 = ",betahat2,"\n")
cat("      -2LL at this parameter value = ",unrest,"\n")
cat("\n")
cat("\n")
cat("Restricted MLE: \n")
cat("  alphahathat = ",alphahathat,"\n")
cat("  betahathat1 = ",betahathat1," betahathat2 = ",betahathat2,"\n")
cat("      -2LL at this parameter value = ",rest,"\n")
cat("\n")
cat("\n")
# Test statistic & p-value
G <- rest-unrest ; p <- 1-pchisq(G,1)
cat("G = ",G,"  p = ",p,"\n")
cat("\n")
cat("\n")
##### End of program gammatest.R #####

```

Here is the output.

```
> source("gammatest.R")
```

```
Unrestricted MLE:
```

```
  alphahat1 = 1.204890  betahat1 = 2.693892  
  alphahat2 = 1.631417  betahat2 = 3.544673  
  -2LL at this parameter value = 620.5435
```

```
Restricted MLE:
```

```
  alphahat1 = 1.426550  
  betahat1 = 2.27531  betahat2 = 4.053725  
  -2LL at this parameter value = 622.253
```

```
G = 1.709518  p = 0.1910479
```

```
Warning messages:
```

```
1: NA/Inf replaced by maximum positive value  
2: NaNs produced in: log(x)  
3: NA/Inf replaced by maximum positive value
```

Just for reference, this example had  $n_1 = 50$ ,  $n_2 = 75$ , and true parameter values  $\alpha_1 = 1$ ,  $\alpha_2 = 1.5$ ,  $\beta_1 = 3$ , and  $\beta_2 = 4$ . The null hypothesis was not rejected at the  $\alpha = 0.05$  level. The results were not statistically significant. All we can say is that these data do not provide sufficient evidence to conclude that  $\alpha_1$  is different from  $\alpha_2$ . Our analysis has led to a Type II error, but that's life. The difference between  $\alpha_1$  and  $\alpha_2$  was too small for us to have a good chance of detecting it with these sample sizes. With bigger sample sizes, the power (probability of rejecting  $H_0$ ) would have been higher.