Likelihood ratio tests

Math 483

1 LR tests for one parameter

Likelihood methods are useful for testing hypotheses, for example

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

The inference proceeds as follows:

1. Compute the likelihood, according to a chosen model $f(x;\theta)$, with the value of $L(\theta_0)$.

2. Compute the MLE $\hat{\theta}$, and its likelihood, $L(\hat{\theta})$. Find *likelihood ratio* $\boxed{\Lambda = L(\hat{\theta})/L(\theta_0)}$. Note that $\Lambda \geq 1$.

3. Let $\lambda \equiv 2\log\Lambda = 2(\log L(\hat{\theta}) - \log L(\theta_0))$. For large sample sizes, under null hypothesis H_0, λ follows chi-square distribution with one degree of freedom.

Explanation: suppose we estimate the mean of a Normal distribution $\theta = \mu$, with known variance σ^2. The MLE is known to be $\overline{X}$. Then, from the Student's theorem,

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \tag{1}$$

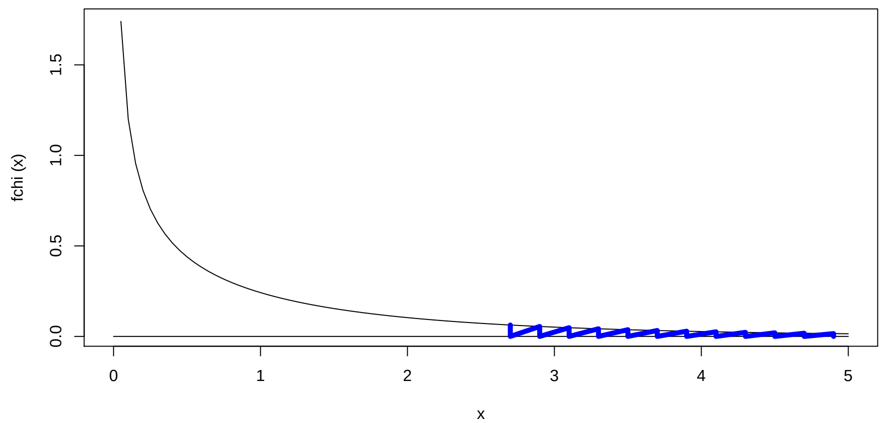
is $\mathcal{N}(0,1)$. On the other hand,

$$2(\log L(\overline{X}) - \log L(\mu)) = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - \sum_{i=1}^{n}(X_i - \overline{X})^2\right].$$

After some algebra, it simplifies to $\frac{n}{\sigma^2}\left(\overline{X} - \mu\right)^2$. But this is exactly the square of (1), and is therefore $\chi^2(1)$.

In general, we can quote Theorem 10.22 on p. 164 from the textbook.
The proof of Theorem 10.22 depends on the ability to approximate log likelihood with an upside-down parabola, and therefore on the finite value of Fisher information $I(\theta)$. Thus, the regularity assumptions are necessary.

(d) We will **reject** H_0 in favor of H_1, whenever $\lambda > \chi^2_{1-\alpha}(1)$, where $\chi^2_{1-\alpha}$ is the $1-\alpha$ quantile of chi-square distribution (`R` command `qchisq(1-`α`,1)`).
($\chi^2_{0.9} \approx 2.7$ is shown)



Example: Suppose $X_1, ..., X_n$ are i.i.d. from $Poisson(\theta)$. Find

$$\Lambda = \prod_{i=1}^{n} e^{-\theta_0+\hat{\theta}} \left(\frac{\theta_0}{\hat{\theta}}\right)^{X_i},$$

$$-2\log\Lambda = 2n(\theta_0 - \hat{\theta}) + 2(\log\hat{\theta} - \log\theta_0)\sum X_i \tag{2}$$

Recall that the MLE for Poisson is again $\overline{X}$. Thus, the expression (2) simplifies to

$$-2\log\Lambda = 2n[(\theta_0 - \overline{X}) + (\log\overline{X} - \log\theta_0)\overline{X}].$$

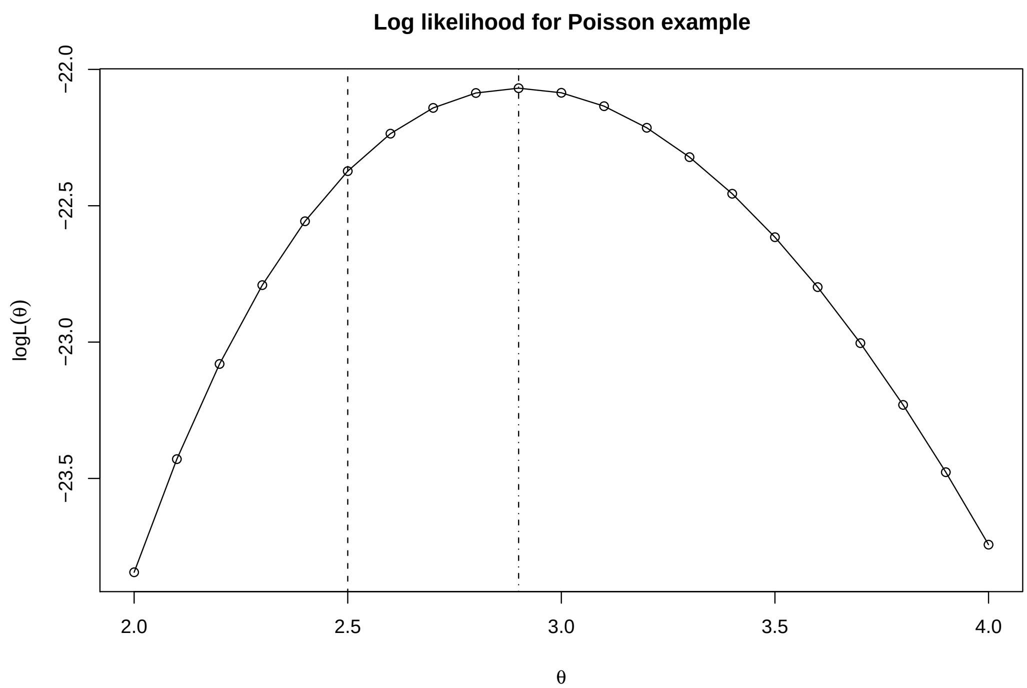
Example data: suppose that the average number of equipment failures in the past has been known to be $\theta_0 = 2.5$ failures per week. After introducing new equipment, we observed number of failures for 10 randomly chosen weeks and obtained the sample $0, 7, 5, 4, 3, 2, 0, 1, 3, 4$. Assume the Poisson model for number of failures. The sample mean is $\overline{X} = 2.9$ (later we'll see that $\overline{X}$ is a *sufficient statistic* for this model, that is, we only need to know $\overline{X}$ to draw our inferences).

Set the hypotheses

$H_0: \theta = 2.5$ "`average number of failures remained the same`"

$H_1: \theta \neq 2.5$ "`average number of failures has changed`"

We then compute $\lambda = -2\log\Lambda = 0.608$. If we used the critical region approach, we'd reject H_0 at the level $\alpha = 0.05$ whenever $\lambda > \chi^2_{0.95}(1) = 3.84$. Thus, we accept H_0(no change).
If we use the p-value approach, we can find p-value$=P(\chi^2 > \lambda) = 1 - \texttt{pchisq}(0.608, 1) = 0.4355$.



Looking at the plot, the difference in heights between the two values is not large enough to be significant.
Question: would we have rejected $H_0 : \theta = 2.0$ based on this data?

```
R code:

X <- c(0, 7, 5, 4, 3, 2, 0, 1, 3, 4)
lpois <- function(th){
     sum(log(dpois(X,th)))
}
xc <- seq(2,4, 0.1)
yc <- xc
nx <- length(xc)
for (i in 1:nx){
  yc[i] <- lpois(xc[i])
}
plot(xc,yc, type="o", main="Log likelihood for Poisson example", xlab=expression(theta),
   ylab = expression(logL(theta)))
 lines(c(2.9,2.9), c(-30,-20), lty = 4)
 lines(c(2.5,2.5), c(-30,-20), lty = 2)
```

2 Multi-dimensional case

Suppose now that the likelihood depends on k parameters, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. Let the null hypothesis be given in terms of q independent constraints:

$$H_0 : g_1(\boldsymbol{\theta}) = a_1, ..., g_q(\boldsymbol{\theta}) = a_q$$

$$H_1 : \text{not all} \quad g_j(\boldsymbol{\theta}) = a_j$$

As an example, consider a goodness-of-fit test (Section 10.4). Let $\theta_1, ..., \theta_k$ be proportions of observations in each of k categories. This will correspond to a classification table with k cells, with the constraint $\sum_{j=1}^{k} \theta_j = 1$.
Then, the null hypothesis that all of the proportions are equal can be expressed as

$$H_0 : \theta_1 = p_{01}, \ ... \theta_k = p_{0k}, \tag{3}$$

that is, for functions $g_j(\boldsymbol{\theta}) = \theta_j$. However, the additional constraint $\sum_{j=1}^{k} \theta_j = 1$ makes one of the equations in H_0 redundant. Thus, here $q = k - 1$.

Now, consider *constrained likelihood*, that is, find $\boldsymbol{\theta}_0$ that solves

$$\text{maximize} \quad l(\boldsymbol{\theta})$$

$$\text{subject to} \quad g_1(\boldsymbol{\theta}) = a_1, ..., g_q(\boldsymbol{\theta}) = a_q$$

The procedure of testing H_0 is then:

- Compute constrained $\boldsymbol{\theta}_0$ and unconstrained MLE $\hat{\boldsymbol{\theta}}$.
- Consider likelihood ratio $\Lambda = L(\boldsymbol{\theta}_0)/L(\hat{\boldsymbol{\theta}})$ (it's always ≤ 1).
- $\lambda = -2 \log \Lambda$ has, for large n, approximate chi-square distribution with q degrees of freedom.
- Reject H_0 at the level α whenever $\lambda > \chi^2_{1-\alpha}(q)$. The p-value$=P(\chi^2 > \lambda) = 1 - \texttt{pchisq}(\lambda, \texttt{q})$.