24th March '25

Statistics

Lecture 15

MATH 414

Vivek T.

Theorem Under suitable conditions on $f(x;\theta)$, any consistent sequence of solutions of mle equations satisfies

$$\boxed{\sqrt{n}\,(\hat{\theta}-\theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)}$$

You also get

$$\sqrt{n I(\theta_0)}\,(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, 1)$$

and

$$\left(\sqrt{n I(\theta_0)}\,(\hat{\theta}_n - \theta_0)\right)^2 \xrightarrow{D} \chi^2(1)$$

You also get

$$\sqrt{n I(\theta_0)}\,(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, 1)$$

and

$$\left(\sqrt{n I(\theta_0)}\,(\hat{\theta}_n - \theta_0)\right)^2 \xrightarrow{D} \chi^2(1)$$

and then if we are ready to believe that

$$I(\hat{\theta}_n) \xrightarrow{P} I(\theta_0)$$

on account of $\hat{\theta}_n \xrightarrow{P} \theta_0$, we have

$$\left(\sqrt{n I(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0)\right)^2 \xrightarrow{D} \chi^2(1)$$

Numerical Implementation of mle in R.

We have seen examples where mle cannot be computed analytically and one must use numerical methods for it.

Example.

Numerical Implementation of mle in R.

We have seen examples where mle cannot be computed analytically and one must use numerical methods for it.

Example. Beta distributions are useful for distributions of proportions.

Here is a list of batting averages from baseball

0.276	0.281	0.225	0.283	0.257
0.250	0.250	0.261	0.312	0.259
0.273	0.222	0.314	0.271	0.294
0.268				

Here is a list of batting averages from baseball

0.276	0.281	0.225	0.283	0.257
0.250	0.250	0.261	0.312	0.259
0.273	0.222	0.314	0.271	0.294
0.268				

We can use max. likelihood estimation to fit a beta distribution to this data.

1

Enter the data as a vector

```
ba <- c(0.276, 0.281, 0.225, 0.283, 0.257,
        0.250, 0.250, 0.261, 0.312, 0.259,
        0.273, 0.222, 0.314, 0.271, 0.294,
        0.268)
```

1. Enter the data as a vector

```
ba <- c( 0.276, 0.281, 0.225, 0.283, 0.257,
         0.250, 0.250, 0.261, 0.312, 0.259,
         0.273, 0.222, 0.314, 0.271, 0.294,
         0.268 )
```

2. Define the log likelihood function

```
loglik <- function (theta, x)
{  if (any (theta <= 0))
      NA
   else sum ( dbeta (x, theta[1], theta[2], log=TRUE))
}
```

③ Use the maxLik package to compute

```
require(maxLik)
m1 <- maxLik(loglik, start
        = c(shape1=1, shape2=1),
        x=ba)

m1
coef(m1)
```

Likelihood Ratio tests

In hypothesis testing with the hypotheses

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

we will create a test based on the likelihood ratio statistic

$$\boxed{\lambda = \frac{L(\theta_0; x)}{L(\hat{\theta}; x)}}$$

This ratio tells us how much more likely is the data when the parameter is equal to the MLE as compared to when the parameter is equal to the null hypothesis value.

We will decide to reject the null hypothesis when λ is small — below a certain threshold.

Example 1

Let $\underline{X} \sim$ Normal(μ, σ^2) be an iid sample. We will use the likelihood ratio to test the null hypothesis $\mu = \mu_0$ if the standard deviation σ is known.

Step 1

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

Step 1

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

Step 2

$$L(\mu) = k_0 \prod_{i=1}^{n} e^{-1/2\sigma^2 (x_i-\mu)^2}$$

$$= k_0 e^{-1/2\sigma^2 \Sigma (x_i-\mu)^2}$$

$$= k_0 e^{-\frac{1}{2\sigma^2} (\Sigma x_i^2 - 2\bar{x}\mu + \mu^2)}$$

Step 1

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

Step 2

$$L(\mu) = k_0 \prod_{i=1}^{n} e^{-1/2\sigma^2 (x_i-\mu)^2}$$

$$= k_0 e^{-1/2\sigma^2 \sum (x_i-\mu)^2}$$

$$= k_0 e^{-\frac{1}{2\sigma^2} (\sum x_i^2 - 2\bar{x}\mu + \mu^2)}$$

$$= k_1 e^{-1/2\sigma^2 (\sum_{i=1}^{n} \bar{x}^2 - 2\bar{x}\mu + \mu^2)}$$

$$= k_1 e^{-\frac{1}{2\sigma^2} n(\bar{x}-\mu)^2} = k_1 e^{-1/2 \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$

$$= k_1 \tilde{L}(\mu)$$

Therefore $\lambda = \dfrac{L(\mu)}{L(\bar{x})} = \dfrac{\tilde{L}(\mu)}{\tilde{L}(\bar{x})}$

$$= \frac{\tilde{L}(\mu)}{1}$$

$$= \tilde{L}(\mu)$$

$$= e^{-1/2\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$

Therefore
$$\lambda = \frac{L(\mu)}{L(\bar{x})} = \frac{\tilde{L}(\mu)}{\tilde{L}(\bar{x})}$$
$$= \frac{\tilde{L}(\mu)}{1}$$
$$= \tilde{L}(\mu)$$
$$= e^{-1/2\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$

Therefore,
$$-2 \log \frac{L(\mu)}{L(\bar{x})} = \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2$$

Therefore $$\lambda = \frac{L(\mu)}{L(\bar{x})} = \frac{\tilde{L}(\mu)}{\tilde{L}(\bar{x})}$$

$$= \frac{\tilde{L}(\mu)}{1}$$

$$= \tilde{L}(\mu)$$

$$= e^{-1/2\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}$$

Therefore,

$$\boxed{W = -2\log\frac{L(\mu)}{L(\bar{x})}} = \left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

Based on this, for a normally distributed sample, we can get an exact test:

Reject H_0 in favour of H_1 if

$$\boxed{\chi_L^2 = -2 \log \lambda \geq \chi_\alpha^2(1)}$$

at the significance level α.

Theorem Assume some regularity conditions on f. Under the null hypothesis,

$H_0: \theta = \theta_0$,

$$\boxed{-2 \log \Lambda \xrightarrow{D} \chi^2(1)}$$

where $\Lambda = \dfrac{L(\theta_0)}{L(\hat{\theta})}$

Based on this, for a large sample, we can get an approximate test:

Reject H_0 in favour of H_1 if

$$\boxed{\chi^2_L = -2\log\lambda \geq \chi^2_\alpha(1)}$$

at the significance level α.

As we mentioned earlier,

$$\left\{ \sqrt{nI(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2 \xrightarrow{D} \chi^2(1)$$

This gives us another test statistic.

$$\chi^2_W = \left\{ \sqrt{nI(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2$$

As we mentioned earlier,

$$\left\{ \sqrt{n I(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2 \xrightarrow{D} \chi^2(1)$$

This gives us another test statistic.

$$\chi^2_W = \left\{ \sqrt{n I(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2$$

Reject H_0 in favour of H_1 if

$$\chi^2_W \geq \chi^2_\alpha(1)$$

As we mentioned earlier,

$$\left\{ \sqrt{nI(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2 \xrightarrow{D} \chi^2(1)$$

This gives us another test statistic.

$$\chi^2_W = \left\{ \sqrt{nI(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0) \right\}^2$$

Reject H_0 in favour of H_1 if

$$\chi^2_W \geq \chi^2_\alpha(1)$$

This is called a <u>Wald-type test</u>

Finally $\quad \dfrac{l'(\theta_0)}{n} \xrightarrow{D} N\left(0, \dfrac{I(\theta_0)}{n}\right)$

or $\quad \dfrac{l'(\theta_0)}{\sqrt{n I(\theta_0)}} \xrightarrow{D} N(0,1)$

or $\quad \left(\dfrac{l'(\theta_0)}{\sqrt{n I(\theta_0)}}\right)^2 \xrightarrow{D} \chi^2(1)$

Finally $$\frac{\ell'(\theta_0)}{n} \xrightarrow{D} N\left(0, \frac{I(\theta_0)}{n}\right)$$

or $$\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}} \xrightarrow{D} N(0,1)$$

or $$\left(\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}}\right)^2 \xrightarrow{D} \chi^2(1)$$

This suggests the test statistic

$$\chi_R^2 = \left(\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}}\right)^2$$

Reject H_0 in favour of H_1 if $\chi_R^2 \geq \chi_\alpha^2(1)$.

Finally $\frac{\ell'(\theta_0)}{n} \xrightarrow{D} N\left(0, \frac{I(\theta_0)}{n}\right)$

or $\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}} \xrightarrow{D} N(0,1)$

or $\left(\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}}\right)^2 \xrightarrow{D} \chi^2(1)$

This suggests the test statistic

$$\chi_R^2 = \left(\frac{\ell'(\theta_0)}{\sqrt{n I(\theta_0)}}\right)^2$$

This is called a Rao-type test.

Reject H_0 in favour of H_1 if $\chi_R^2 \geq \chi_\alpha^2(1)$.

Example. Let $X_1, X_2, \ldots, X_n$ be a random sample having the common beta$(1, \theta)$ p.d.f.

Consider the hypotheses

$$H_0: \theta_0 = 1 \quad \text{versus} \quad H_1: \theta_0 \neq 1$$

Recall that $$\hat{\theta} =_n \frac{-n}{\sum_{i=1} \log X_i}$$

Therefore $$L(\hat{\theta}) = \prod_{i=1}^{n} \hat{\theta} (X_i)^{\hat{\theta}-1}$$

$$l(\hat{\theta}) = \sum_{i=1}^{n} \left\{ \log \hat{\theta} + (\hat{\theta}-1) \log X_i \right\}$$

$$= n \log \hat{\theta} + (\hat{\theta}-1) \sum \log X_i$$

$$= n \log n - n \log \left(-\sum_{i=1}^{n} \log X_i\right)$$

$$- n - \sum_{i=1}^{n} \log X_i$$

$$= n \log(n-1) - n \log \left(-\sum_{i=1}^{n} \log X_i\right)$$

$$- \sum_{i=1}^{n} \log X_i$$

$$\ell(\hat{\theta}) = \sum_{i=1}^{n} \left\{ \log \hat{\theta} + (\hat{\theta}-1) \log X_i \right\}$$

$$= n \log \hat{\theta} + (\hat{\theta}-1) \sum \log X_i$$

$$= n \log n - n \log \left(-\sum_{i=1}^{n} \log X_i\right) - n - \sum_{i=1}^{n} \log X_i$$

$$= n \log (n-1) - n \log \left(-\sum_{i=1}^{n} \log X_i\right) - \sum_{i=1}^{n} \log X_i$$

$$L(1) = 1 \quad \Rightarrow \quad \ell(1) = 0$$

Therefore, $\chi_2^2 = -2 \log \Lambda = 2\left(n \log (n-1) - \sum_{i=1}^{n} \log X_i - n \log \left(-\sum_{i=1}^{n} \log X_i\right)\right)$

Next, let us look at the <u>Wald test statistic</u>.

Next, let us look at the Wald test statistic.

$$\chi^2_W = \left(\sqrt{nI(\hat{\theta}_n)}\,(\hat{\theta}_n - \theta_0)\right)^2$$

$$= \left\{\sqrt{n \cdot \frac{1}{\hat{\theta}_n^2}} \cdot (\hat{\theta} - 1)\right\}^2$$

$$= \frac{n}{\hat{\theta}^2} \cdot (\hat{\theta} - 1)^2 = n\left(1 - \frac{1}{\hat{\theta}}\right)^2$$

$$= \frac{\left(\Sigma \log x_i + n\right)^2}{n}$$

And finally the Rao type test statistic

$$\chi_R^2 = \left(\frac{l'(\theta_0)}{\sqrt{n I(\theta_0)}} \right)^2$$

$$\boxed{l'(\theta) = \sum_{i=1}^{n} \log x_i + \frac{n}{\theta}}$$

$(\theta_0 = 1)$

$$= \left(\frac{\sum \log x_i + n}{\sqrt{n}} \right)^2$$

$$\left(\because\ I(\theta_0) = \frac{1}{\theta_0^2} = 1 \right)$$

And finally the Rao type test statistic

$$\chi_R^2 = \left(\frac{l'(\theta_0)}{\sqrt{n I(\theta_0)}} \right)^2$$

$$\boxed{l'(\theta) = \sum_{i=1}^{n} \log x_i + \frac{n}{\theta}}$$

$(\theta_0 = 1)$

$$= \left(\frac{\sum \log x_i + n}{\sqrt{n}} \right)^2$$

$$\left(\because \; I(\theta_0) = \frac{1}{\theta_0^2} = 1 \right)$$

Same as χ_W^2.

An example. An instructor believes that the number of students who arrive late for class should follow a Poisson distribution.

x	0	1	2	3	4
frequency	3	3	2	1	1

$$\text{mle of } \lambda = \bar{X} = \frac{3+4+3+4}{10}$$

$$= 1.4.$$

Suppose we want to test the hypotheses

$H_0: \lambda = 1$, $H_a: \lambda \neq 1$.

In this case, the likelihood ratio statistic is

$$\lambda = \frac{L(1;x)}{L(\bar{x};x)} = \prod \frac{(e^{-1} 1^{x})/x!}{\frac{e^{-\bar{x}} \bar{x}^{x}}{x!}}$$

$$= \prod \frac{e^{-1}}{e^{-\bar{x}} \bar{x}^{x}}.$$

Therefore,

$$-2 \ln \lambda = 2 \sum_{x} \{-\bar{x} + x \log \bar{x} + 1\}$$

$$= 2[-n\bar{x} + n\bar{x} \log \bar{x} + n]$$

From the data,

$$-2\ln\lambda = 1.4212.$$

Now, we can estimate the p-value using χ^2 approximation.

```
x <- c(1, 1, 0, 4, 2, 1, 3, 0, 0, 2);
tally(x)
mean(x)
lr <- 2*(-n * mean(x) + n * mean(x) *
            log(mean(x)) + n]
pval <- 1- pchisq(lr, df = 1)
```

You can also use

simulations.