Topic 15
Maximum Likelihood Estimation
Examples and Asymptotic Properties

Outline

Normal Random Variables

For a simple random sample of n normal random variables, we can use the properties of the exponential function to simplify the likelihood function.

$$\begin{aligned}\mathbf{L}(\mu,\sigma^2|\mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x_1-\mu)^2}{2\sigma^2}\right)\cdots\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(x_n-\mu)^2}{2\sigma^2}\right)\\ &= \frac{1}{\sqrt{(2\pi\sigma^2)^n}}\exp-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.\end{aligned}$$

The log-likelihood $\ln\mathbf{L}(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}(\ln 2\pi + \ln\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$.

The score function is now a vector $\left(\frac{\partial}{\partial\mu}\ln\mathbf{L}(\mu,\sigma^2|\mathbf{x}), \frac{\partial}{\partial\sigma^2}\ln\mathbf{L}(\mu,\sigma^2|\mathbf{x})\right)$. Next we find the zeros to determine the maximum likelihood estimators $\hat{\mu}$ and $\hat{\sigma}^2$.

Normal Random Variables

$$\ln \mathbf{L}(\mu, \sigma^2|\mathbf{x}) = -\frac{n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$0 = \frac{\partial}{\partial \mu} \ln \mathbf{L}(\hat{\mu}, \hat{\sigma}^2|\mathbf{x}) = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu}) = \frac{1}{\hat{\sigma}^2} n(\bar{x} - \hat{\mu}).$$

Because the second partial derivative with respect to μ is negative, $\hat{\mu}(\mathbf{x}) = \bar{x}$ is the maximum likelihood estimator. For the derivative with respect to σ^2,

$$0 = \frac{\partial}{\partial \sigma^2} \ln \mathbf{L}(\hat{\mu}, \hat{\sigma}^2|\mathbf{x}) = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = -\frac{n}{2(\hat{\sigma}^2)^2}\left(\hat{\sigma}^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2\right).$$

Recalling that $\hat{\mu}(\mathbf{x}) = \bar{x}$, we obtain a *biased estimator*,

$$\hat{\sigma}^2(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Mark and Recapture

We return to consider Lincoln-Peterson method of mark and recapture and find its maximum likelihood estimate. Recall that

- t be the number captured and tagged,
- k be the number in the second capture,
- r be the number in the second capture that are tagged, and let
- N be the total population size.

Thus, t and k is under the control of the experimenter. The value of r is random and the populations size N is the parameter to be estimated.

Mark and Recapture

The likelihood function for N is the hypergeometric distribution

$$L(N|r) = \frac{\binom{t}{r}\binom{N-t}{k-r}}{\binom{N}{k}}.$$

Exercise. Show that the maximum likelihood estimate

$$\hat{N} = \left[\frac{tk}{r}\right].$$

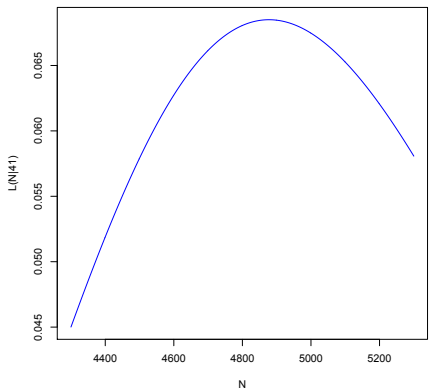
where $[\cdot]$ mean the greatest integer less than.
Hint: Find the values of N for which $L(N|r)/L(N-1|r) > 1$.

Thus, the maximum likelihood estimate is, in this case, obtained from the method of moments estimate by rounding down to the next integer.

Mark and Recapture

We return to the simulation of a lake having 4500 fish.

```
> N<-4500;t<-400;k<-500
> fish<-c(rep(1,t),rep(0,N-t))
> (r<-sum(sample(fish,k)))
[1] 41
> (Nhat<-floor(k*t/r))
[1] 4878
> N<-c(4300:5300)
> L<-dhyper(r,t,N-t,k)
> plot(N,L,type="l",
  ylab="L(N|41)",col="blue")
```



Plot of likelihood from the simulation with $r = 41$. The maximum $\hat{N} = 4878$.

Linear Regression

Our data are n observations. The responses y_i are linearly related to the explanatory variable x_i with an error ϵ_i,

$$y_i = \alpha + \beta x_i + \epsilon_i.$$

Here we take the ϵ_i to be independent $N(0, \sigma)$ random variables. Our model has three parameters, the intercept α, the slope β, and the variance of the error σ^2.
Thus, the joint density for the ϵ_i is

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{\epsilon_1^2}{2\sigma^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{\epsilon_2^2}{2\sigma^2} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{\epsilon_n^2}{2\sigma^2} = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \epsilon_i^2$$

Since $\epsilon_i = y_i - (\alpha + \beta x_i)$, the likelihood function,

$$L(\alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2.$$

Linear Regression

The logarithm

$$\ln L(\alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x}) = -\frac{n}{2}(\ln 2\pi + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2.$$

Consequently, maximizing the likelihood function for the parameters α and β is equivalent to minimizing

$$SS(\alpha.\beta) = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2.$$

The principle of maximum likelihood is equivalent to the least squares criterion. Thus,

$$\hat{\beta} = \frac{\text{cov}(x, y)}{\text{var}(x)}, \quad \text{and} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Linear Regression

Exercise. Show that the maximum likelihood estimator for σ^2 is

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{k=1}^{n}(y_i - \hat{y}_i)^2.$$

where $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$. are the predicted values from the regression line.

Frequently, software will report the unbiased estimator. For ordinary least square procedures, this is

$$\hat{\sigma}^2_{U} = \frac{1}{n-2}\sum_{k=1}^{n}(y_i - \hat{y}_i)^2.$$

For the measurements on the lengths in centimeters of the femur and humerus for the five specimens of *Archeopteryx*, we have the following R output for linear regression.

Linear regression

```
> femur<-c(38,56,59,64,74), humerus<-c(41,63,70,72,84)
> summary(lm(humerus~femur))
Call:
lm(formula = humerus ~ femur)
Residuals:
      1       2       3       4       5
-0.8226 -0.3668  3.0425 -0.9420 -0.9110
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.65959    4.45896  -0.821 0.471944
femur        1.19690    0.07509  15.941 0.000537 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 1.982 on 3 degrees of freedom
Multiple R-squared: 0.9883, Adjusted R-squared: 0.9844
F-statistic: 254.1 on 1 and 3 DF,  p-value: 0.0005368
```

Asymptotic Properties

Much of the attraction of maximum likelihood estimators is based on their properties for large sample sizes.

1. Consistency. If θ_0 is the state of nature and $\hat{\theta}_n(X)$ is the maximum likelihood estimator based on n observations from a simple random sample, then

$$\hat{\theta}_n(X) \to \theta_0 \quad \text{as } n \to \infty.$$

In words, as the number of observations increase, the distribution of the maximum likelihood estimator becomes more and more concentrated about the true state of nature.

Asymptotic Properties

2. Asymptotic normality and efficiency. Under some technical assumptions

$$\sqrt{n}(\hat{\theta}_n(X) - \theta_0).$$

converges in distribution as $n \to \infty$ to a normal random variable with mean 0 and variance $1/I(\theta_0)$, the Fisher information for one observation. Thus,

$$\text{Var}_{\theta_0}(\hat{\theta}_n(X)) \approx \frac{1}{nI(\theta_0)},$$

the lowest variance possible under the Crámer-Rao lower bound. Let

$$Z_n = \frac{\hat{\theta}(X) - \theta_0}{1/\sqrt{nI(\theta_0)}}.$$

Then, as with the central limit theorem, Z_n converges in distribution to a standard normal random variable.

Asymptotic Properties

3. Properties of the log likelihood surface. For large sample sizes, the variance of a maximum likelihood estimator is approximately the reciprocal of the Fisher information

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln L(\theta|X)\right].$$

The Fisher information can be approximated by the observed information based on the data $\mathbf{x}$,

$$J(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2}\ln L(\hat{\theta}(\mathbf{x})|\mathbf{x}),$$

giving the negative of the curvature of the log-likelihood surface at the maximum likelihood estimate $\hat{\theta}(\mathbf{x})$.

- If the curvature is small near the maximum likelihood estimator, then the likelihood surface is nearty flat and the variance is large.
- If the curvature is large, the likelihood decreases quickly at the maximum and thus the variance is small.