Tutorial: Statistical Inference

Hongyang Li, Francis

April 13, 2016

The Chinese University of Hong Kong
Recap
Terminology on Estimators

Let \( \hat{\Theta}_n \) be an estimator of an unknown parameter \( \theta \), that is, a function of \( n \) observations \( X_1, \ldots, X_n \) whose distribution depends on \( \theta \).

- The estimation error \( \tilde{\Theta}_n = \hat{\Theta}_n - \theta \).
- The bias of the estimator is the expected value of the estimation error. \( b_\theta(\hat{\Theta}_n) = \mathbb{E}[\hat{\Theta}_n] - \theta \).
- We call \( \hat{\Theta}_n \) unbiased if \( \mathbb{E}[\hat{\Theta}_n] = \theta, \forall \theta \).
- We call \( \hat{\Theta}_n \) asymptotically unbiased if \( \lim_{n \to \infty} \mathbb{E}[\hat{\Theta}_n] = \theta, \forall \theta \).
- We call \( \hat{\Theta}_n \) consistent if the sequence \( \hat{\Theta}_n \) converges to the true value of \( \theta, \forall \theta \).
We are given the realization $x = (x_1, \ldots, x_n)$ of a random vector $X = (X_1, \ldots, X_n)$, distributed according to a PMF $p_X(x; \theta)$ or PDF $f_X(x; \theta)$.

The maximum likelihood (ML) estimate is a value of $\theta$ that maximizes the likelihood function, $p_X(x; \theta)$ or $f_X(x; \theta)$, over all $\theta$.

The ML estimate of a one-to-one function $h(\theta)$ is $h(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the ML estimate of $\theta$ (the invariance principle).

When the random variables $X_i$ are i.i.d., and under some mild additional assumptions, each component of the ML estimator is consistent and asymptotically normal.
Estimates of Mean and Variance

Let the observations $X_1, \ldots, X_n$ be i.i.d., with mean $\theta$ and variance $\nu$ that are unknown.

- The sample mean
  
  $$M_n = \frac{X_1 + \cdots + X_n}{n}$$

  is an unbiased estimator of $\theta$, and its mean squared error is $\frac{\nu}{n}$.

- Two variance estimators are
  
  $$\bar{S}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_n)^2, \quad \hat{S}_n^2 = \frac{1}{n - 1} \sum_{i=1}^{n} (X_i - M_n)^2.$$

- The estimator $\bar{S}_n^2$ coincides with the ML estimator if the $X_i$ are normal. It is biased but asymptotically unbiased. The estimator $\hat{S}_n^2$ is unbiased. For large $n$, the two variance estimators essentially coincide.
A confidence interval for a scalar unknown parameter $\theta$ is an interval whose endpoints $\hat{\Theta}_n^-$ and $\hat{\Theta}_n^+$ bracket $\theta$ with a given high probability.

$\hat{\Theta}_n^-$ and $\hat{\Theta}_n^+$ are random variables that depend on the observations $X_1, \ldots, X_n$.

A $1 - \alpha$ confidence interval is one that satisfies

$\mathbb{P}_\theta(\hat{\Theta}_n^- \leq \theta \leq \hat{\Theta}_n^+) \geq 1 - \alpha$, for all possible value of $\theta$. 
Given $n$ data pairs $(x_i, y_i)$, the estimates that minimize the sum of the squared residuals are given by

$$
\hat{\theta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}
$$

where $\bar{x}, \bar{y}$ are the average of samples.

Extension: From 1-D to N-D application.
Examples
Example 1

The leaning tower of Pisa continuously tilts over time. Measurements between years 1975 and 1987 of the "lean" of a fixed point on the tower (the distance in meters of the actual position of the point, and its position if the tower were straight) have produced the following table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean</td>
<td>2.9642</td>
<td>2.9644</td>
<td>2.9656</td>
<td>2.9667</td>
<td>2.9673</td>
<td>2.9688</td>
<td>2.9696</td>
</tr>
<tr>
<td>Lean</td>
<td>2.9698</td>
<td>2.9713</td>
<td>2.9717</td>
<td>2.9725</td>
<td>2.9742</td>
<td>2.9757</td>
<td></td>
</tr>
</tbody>
</table>
Solution to Example 1

Using the regression formulas, we obtain

\[
\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0.0009, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x} = 1.1233,
\]

where \( \bar{x} = 1981, \bar{y} = 2.9694 \). The estimated linear model is \( y = 0.0009x + 1.1233 \).
Consider the polling problem of Section 5.4 (Example 5.11), where we wish to estimate the fraction $\theta$ of voters who support a particular candidate for office. We collect $n$ independent sample voter responses $X_1, \ldots, X_n$, where $X_i$ is viewed as a Bernoulli random variable, with $X_i = 1$ if the $i$-th voter supports the candidate.

**Question:** Estimate $\theta$ with the sample mean $\hat{\Theta}_n$, and construct a confidence interval based on a normal approximation and different ways of estimating or approximating the unknown variance.

For concreteness, suppose that 684 out of a sample of $n = 1200$ voters support the candidate, so that $\hat{\Theta}_n = 684/1200 = 0.57$. 
(a) If we use the variance estimate

\[
\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\Theta}_n)^2 \\
= \frac{1}{1199} \left( 684 \cdot \left(1 - \frac{684}{1200}\right)^2 + (1200 - 684) \cdot \left(0 - \frac{684}{1200}\right)^2 \right) \\
\approx 0.245,
\]

and treat \( \hat{\Theta}_n \) as a normal random variable with mean \( \theta \) and variance 0.245, we obtain the 95% confidence interval as follows:

\[
\left[ \hat{\Theta}_n - 1.96 \frac{\hat{S}_n}{\sqrt{n}}, \hat{\Theta}_n + 1.96 \frac{\hat{S}_n}{\sqrt{n}} \right] = \left[ 0.57 - \frac{1.96 \cdot \sqrt{0.245}}{\sqrt{1200}}, 0.57 + \frac{1.96 \cdot \sqrt{0.245}}{\sqrt{1200}} \right] \\
= [0.542, 0.598].
\]
Solution to Example 2 (cont’d)

(b) The variance estimate

$$\hat{\Theta}_n(1 - \hat{\Theta}_n) = \frac{684}{1200} \cdot (1 - \frac{684}{1200}) = 0.245$$

is the same as the previous one (up to three decimal place accuracy), and the resulting 95% confidence interval

$$\left[\hat{\Theta}_n - 1.96 \frac{\sqrt{\hat{\Theta}_n(1 - \hat{\Theta}_n)}}{\sqrt{n}}, \hat{\Theta}_n + 1.96 \frac{\sqrt{\hat{\Theta}_n(1 - \hat{\Theta}_n)}}{\sqrt{n}}\right]$$

is again [0.542, 0.598].
(c) The conservative upper bound of $1/4$ for the variance results in the confidence interval

$$\left[ \hat{\Theta}_n - 1.96 \frac{1/2}{\sqrt{n}}, \hat{\Theta}_n + 1.96 \frac{1/2}{\sqrt{n}} \right] = \left[ 0.57 - \frac{1.96 \cdot (1/2)}{\sqrt{1200}}, 0.57 + \frac{1.96 \cdot (1/2)}{\sqrt{1200}} \right]$$

$$= [0.542, 0.599],$$

which is only slightly wider, but practically the same as before.
Example 3

Given the five data pairs \((x_i, y_i)\) in the table below,

<table>
<thead>
<tr>
<th>x</th>
<th>0.8</th>
<th>2.5</th>
<th>5</th>
<th>7.3</th>
<th>9.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>-2.3</td>
<td>20.9</td>
<td>103.5</td>
<td>215.8</td>
<td>334</td>
</tr>
</tbody>
</table>

we want to construct a model relating \(x\) and \(y\). We consider a linear model

\[ Y_i = \theta_0 + \theta_1 x_i + W_i, \quad i = 1, \ldots, 5, \]

and a quadratic model

\[ Y_i = \beta_0 + \beta_1 x_i^2 + V_i, \quad i = 1, \ldots, 5. \]

where \(W_i\) and \(V_i\) represent additive noiseterms, modeled by independent normal random variables with mean zero and variance \(\sigma_1^2\) and \(\sigma_2^2\), respectively.

1. Find the ML estimates of the linear model parameters.
2. Find the ML estimates of the quadratic model parameters.
We can use the regression formulas in a similar way.

(a) 
\[
\hat{\theta}_1 = \frac{\sum_{i=1}^{5}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{5}(x_i - \bar{x})^2} = 40.53, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1\bar{x} = -65.86,
\]
where \(\bar{x} = 4.94, \bar{y} = 134.38\). The estimated linear model is \(y = 40.53x - 65.86\).

(b) 
\[
\hat{\theta}_1 = \frac{\sum_{i=1}^{5}(x_i^2 - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{5}(x_i^2 - \bar{x})^2} = 4.09, \quad \hat{\theta}_0 = \bar{y} - \hat{\theta}_1\bar{x} = -3.07,
\]
where \(\bar{x} = 33.60, \bar{y} = 134.38\). The estimated quadratic model is \(y = 4.09x^2 - 3.07\).
Solution to Example 3 (cont’d)

Sample data points

Estimated first-order model

Estimated second-order model