ECE 645: Estimation Theory
Spring 2015
Instructor: Prof. Stanley H. Chan

Lecture 4: Law of Large Number and Central Limit Theorem

(LaTeX prepared by Jing Li)
March 31, 2015

This lecture note is based on ECE 645(Spring 2015) by Prof. Stanley H. Chan in the School of Electrical and Computer Engineering at Purdue University.

1 Probability Bounds for P_F and P_M

In the previous lectures we have studied various detection methods. Starting from this lecture, we want to take a step further to analyze the performance of these detection methods. In order to motivate ourselves to learn a set of new tools called Large Deviation Theory, let us first review some "standard" tools, namely the Law of Large Number and the Central Limit Theorem.

To begin our discussion, let us first consider the probability of false alarm P_F and the probability of miss P_M. If $Y = y$ is a one-dimensional observation, we can show the following proposition.

Proposition 1.
Given a one-dimensional observation $Y = y$ and a decision rule $\delta(y)$, it holds that

$$P_F \leq \mathbb{P}(\ell(y) \geq \eta | H_0), \tag{1}$$

and

$$P_M \leq \mathbb{P}(\ell(y) \leq \eta | H_1), \tag{2}$$

where $\ell(y) \overset{\text{def}}{=} \log L(y)$ is the log-likelihood ratio.

Proof.
Given $\delta(y)$, it holds that

$$P_F = \int\limits_{\ell(y)>\eta} f_0(y)\,dy + \gamma \int\limits_{\ell(y)=\eta} f_0(y)\,dy \leq \int\limits_{\ell(y)\geq\eta} f_0(y)\,dy = \mathbb{P}(\ell(y) \geq \eta | H_0),$$

where the inequality holds because $\gamma \leq 1$. Similarly, we have

$$P_M = \int\limits_{\ell(y)<\eta} f_1(y)\,dy + (1-\gamma) \int\limits_{\ell(y)=\eta} f_1(y)\,dy \leq \int\limits_{\ell(y)\leq\eta} f_1(y)\,dy = \mathbb{P}(\ell(y) \leq \eta | H_1).$$

$\square$

While the derivation shows that P_F and P_M can be evaluated through the probability of having $\ell(y) \lesseqgtr \eta$, the same trick becomes much more difficult if we proceed to a high-dimensional observation $\boldsymbol{Y} = \boldsymbol{y}$. In this case, we let

$$\boldsymbol{y} = [y_1, y_2, \ldots, y_n]^T. \tag{3}$$

Then,

$$\int\limits_{\ell(\boldsymbol{y})\geq\eta} f_0(\boldsymbol{y})\,d\boldsymbol{y} = \int\limits_{\ell(\boldsymbol{y})\geq\eta} f_0(y_1,\ldots,y_n)\,dy_1\ldots dy_n$$
$$= \int\limits_{\ell(\boldsymbol{y})\geq\eta} \prod_{i=1}^{n} f_0(y_i)\,dy_1\ldots dy_n. \tag{4}$$

Unfortunately, (4) involves multivariate integration and is extremely difficult to compute.

To overcome this difficulty, it will be useful to note that

$$P_F \leq \mathbb{P}(\ell(\boldsymbol{y}) \geq \eta | H_0). \tag{5}$$

Since

$$\ell(\boldsymbol{y}) = \log\frac{f_1(\boldsymbol{y})}{f_0(\boldsymbol{y})} = \sum_{i=1}^{n} \ell_i(y_i),$$

where $\ell_i(y_i) \overset{\text{def}}{=} \log\frac{f_1(y_i)}{f_0(y_i)}$, it holds that

$$\mathbb{P}(\ell(\boldsymbol{y}) \geq \eta | H_0) = \mathbb{P}\left[\sum_{i=1}^{n} \ell_i(y_i) \geq \eta \;\middle|\; H_0\right]. \tag{6}$$

By letting $X_i = \ell_i(y_i)$, we see that P_F can be equivalently bounded as

$$P_F \leq \mathbb{P}\left[\sum_{i=1}^{n} X_i \geq \eta \;\middle|\; H_0\right]. \tag{7}$$

Therefore, if we can derive an accurate **upper bounds** for $\mathbb{P}(\sum_{i=1}^{n} X_i \geq \eta \mid H_0)$, then we can find an upper bound of P_F. So the question now is: How do we find good upper bounds for $\mathbb{P}(\sum_{i=1}^{n} X_i \geq \eta \mid H_0)$?

2 Weak Law of Large Number

We begin the analysis by reviewing some elementary probability inequalities.

Theorem 1. Markov Inequality
For any random variable $X \geq 0$, and for any $\epsilon > 0$,

$$\mathbb{P}(X > \epsilon) \leq \frac{\mathbb{E}[X]}{\epsilon} \tag{8}$$

Proof.

$$\epsilon\mathbb{P}(X > \epsilon) = \epsilon\int_{\epsilon}^{\infty} f_X(x)\,dx \overset{(a)}{\leq} \int_{\epsilon}^{\infty} x f_X(x)\,dx \overset{(b)}{\leq} \int_{0}^{\infty} x f_X(x)\,dx = \mathbb{E}[X],$$

where (a) holds because $\epsilon < x$, and (b) holds because $x f_X(x) \geq 0$.

□

TO DO: Add a pictorial explanation using $\mathbb{E}[X] = \int_0^\infty (1 - F_X(x))dx$.

Theorem 2. CHEBYSHEV INEQUALITY
Let X be a random variable such that $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) < \infty$. Then, for all $\epsilon > 0$,

$$\mathbb{P}(|X - \mu| > \epsilon) \leq \frac{\mathrm{Var}[X]}{\epsilon^2}. \tag{9}$$

Proof.

$$\mathbb{P}(|X - \mu| > \epsilon) = \mathbb{P}((X - \mu)^2 > \epsilon^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\epsilon^2} = \frac{\mathrm{Var}[X]}{\epsilon^2}$$

where the inequality is due to Markov.

□

With Chebyshev inequality, we can now prove the following result.

Proposition 2.
Let $X_1, \ldots, X_n$ be iid random variables with $\mathbb{E}[X_k] = \mu$ and $Var(X_k) = \sigma^2$. If

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

then for any $\epsilon > 0$, we have

$$\mathbb{P}(|Y_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}. \tag{10}$$

Proof.
By Chebyshev inequality, we have

$$\mathbb{P}(|Y_n - \mu| > \epsilon) \leq \frac{\mathbb{E}[(Y_n - \mu)^2]}{\epsilon^2}.$$

Now, we can show that

$$\mathbb{E}[(Y_n - \mu)^2] = \mathrm{Var}(Y_n) = \mathrm{Var}\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var}(X_k) = \frac{\sigma^2}{n}.$$

□

The interpretation of Proposition 2 is important. It says that if we have a sequence of iid random variables $X_1, \ldots, X_n$, the mean Y_n will stay around the mean of X_1. In particular:

$$\lim_{n \to \infty} \mathbb{P}(|Y_n - \mu| > \epsilon) \leq \lim_{n \to \infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

This result is known as the **Weak Law of Large Numbers**.

Example

Consider a unit square containing an arbitrary shape Ω. Let $X_1, \ldots, X_n$ be a sequence of iid Bernoulli random variables with probability $p = |\Omega|$, i.e., $p =$ area of Ω. Let $Y_n = \frac{1}{n}\sum_{k=1}^n X_k$. We can show that

$$\mathbb{E}[Y_n] = \frac{1}{n}\sum_{k=1}^n \mathbb{E}[X_k] = \frac{np}{n} = p, \tag{11}$$

and

$$Var(Y_n) = \frac{1}{n^2}\sum_{k=1}^n \mathrm{Var}(X_k) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}. \tag{12}$$

Therefore:

$$\mathbb{P}(|Y_n - \mu| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \to 0 \text{ as } n \to \infty.$$

So by throwing arbitrarily n "darts" to the unit square we can approximate the area Ω.

Example

TO DO: Add an example of approximating $y = \sum_{i=1}^n a_i x_i$ by $Y = \sum_{i=1}^n a_i x_i I_i / p_i$.

The convergence behavior demonstrated by WLLN is known as the convergence in probability. Formally, it says the followings.

Definition 1. Convergence in Probability
We say that a sequence of random variables $Y_1, \ldots, Y_n$ converges *in probability* to μ, denoted by $Y_n \xrightarrow{p} \mu$ if

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - \mu| > \epsilon) = 0. \tag{13}$$

For more discussion regarding WLLN, we refer the readers to standard probability textbooks.

We close this section by mentioning the following proposition, which appears to be very useful in practice.

Proposition 3.
If $Y_n \xrightarrow{p} \mu$, then $f(Y_n) \xrightarrow{p} f(\mu)$ for any function f that is continuous at μ.

Proof.
Since f is continuous at μ, by continuity we must have that $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$|x - \mu| < \delta \Rightarrow |f(x) - f(\mu)| < \epsilon.$$

Therefore,

$$\mathbb{P}(|Y_n - \mu| < \delta) \leq \mathbb{P}(|f(x) - f(\mu)| < \epsilon)$$

because "$|Y_n - \mu| < \delta$" is a subset of "$|f(x) - f(\mu)| < \epsilon$". Hence

$$\mathbb{P}(|Y_n - \mu| < \delta) \leq \mathbb{P}(|f(x) - f(\mu)| < \epsilon) \to 0 \text{ as } n \to \infty$$

$\square$

Example

Let $X_1, ..., X_n$ be iid Poisson(λ). Then if $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, and $Y_n \xrightarrow{p} \lambda$, then

$$e^{-Y_n} \xrightarrow{p} e^{-\lambda}$$

3 Central Limit Theorem

In introductory probability courses we have also learned the Central Limit Theorem. Central Limit Theorem concerns about the convergence of a sequence of distributions.

Definition 2.
A sequence of distributions with CDF $F_1, \ldots, F_n$ is said to converge to another distribution F, denoted as $F_n \to F$, if $F_n(x) \to F(x)$ at all continuous points x of F.

Definition 3. Convergence in Distribution
A sequence of random variables $Y_1, \ldots, Y_n$ is said to converge to Y in distribution F, denoted as $Y_n \xrightarrow{d} F$, if $F_n \to F$, where F_n is the cdf of Y_n and F is the CDF of Y.

Example

The notation $Y_n \xrightarrow{d} \mathcal{N}(0,1)$ means that the distribution of Y_n is converging to $\mathcal{N}$(0,1). Note that $Y_n \xrightarrow{d} Y$ does not mean that Y_n is becoming Y. It only means that F_{Y_n} is becoming F_Y.

Remark

$Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y$, but the converse is not true. For example, let X and Y be two iid random variables with distribution $\mathcal{N}(0,1)$. Let $Y_n = Y + \frac{1}{n}$. Then it can be shown that $Y_n \xrightarrow{p} Y$, as well as $Y_n \xrightarrow{d} Y$. This gives $Y_n \xrightarrow{d} X$, as X has the same distribution as Y. However $Y_n \xrightarrow{p} X$ is not true, as Y_n is becoming Y, not X.

We now present the Central Limit Theorem.

Theorem 3. Central Limit Theorem
Let $X_1, ..., X_n$ be iid random variables with $\mathbb{E}[X_k] = \mu$ and $\mathrm{Var}(X_k) = \sigma^2 < \infty$, Then

$$\sqrt{n}(Y_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$.

Proof.
It is sufficient to prove that

$$\sqrt{n}\left(\frac{Y_n - \mu}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1)$$

Let $Z_n = \sqrt{n}(\frac{Y_n - \mu}{\sigma})$. The moment generating function of Z_n is

$$M_{Z_n}(s) \stackrel{\text{def}}{=} \mathbb{E}[e^{sZ_n}] = \mathbb{E}\left[e^{s\sqrt{n}(\frac{Y_n-\mu}{\sigma})}\right] = \prod_{k=1}^{n} \mathbb{E}\left[e^{\frac{s}{\sigma\sqrt{n}}(X_k-\mu)}\right].$$

By Taylor approximation, we have

$$\mathbb{E}\left[e^{\frac{s}{\sigma\sqrt{n}}(X_k-\mu)}\right] = \mathbb{E}\left[1+\frac{s}{\sigma\sqrt{n}}(X_k-\mu)+\frac{s^2}{\sigma^2 n}(X_k-\mu)^2+O(\frac{1}{\sigma^3\sqrt{n^3}}(X_k-\mu)^3)\right]$$
$$= (1+0+\frac{s^2}{2n}).$$

Therefore,

$$M_{Z_n}(s) = \left(1+0+\frac{s^2}{2n}\right)^n \xrightarrow{(a)} e^{\frac{s^2}{2}},$$

as $n \to \infty$. To prove (a), we let $y_n = (1+\frac{s^2}{2n})^n$. Then, $\log y_n = n\log(1+\frac{s^2}{2n})$, and by Taylor approximation we have

$$\log(1+x_0) \approx x_0 - \frac{x_0^2}{2}.$$

Therefore,

$$\log y_n = n\log(1+\frac{s^2}{2n}) = n(\frac{s^2}{2n}-\frac{s^4}{4n^2}) = \frac{s^2}{2}-\frac{s^4}{4n} \xrightarrow{n\to\infty} \frac{s^2}{2}.$$

□

As a corollary of the Central Limit Theorem, we also derive the following proposition.

Proposition 4. Delta Method
If $\sqrt{n}(T_n-\theta) \xrightarrow{d} \mathcal{N}(0,\tau^2)$, then $\sqrt{n}(f(T_n)-f(\theta)) \xrightarrow{d} \mathcal{N}(0,\tau^2(f'(\theta)^2))$, provided $f'(\theta)$ exists.

This result is known as the Delta Method.

Proof.
By Taylor expansion

$$f(T_n) = f(\theta) + (T_n-\theta)f'(\theta) + O((T_n-\theta)^2)$$

Therefore,

$$\sqrt{n}(f(T_n)-f(\theta)) = \sqrt{n}(T_n-\theta)f'(\theta) \xrightarrow{d} \mathcal{N}(0,\tau^2(f'(\theta)^2)).$$

□

We close this section by discussing the limitation of the Central Limit Theorem. Recall that our analysis question is to study:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq \eta\right). \tag{14}$$

Central Limit Theorem says that

$$\lim_{n\to\infty} \mathbb{P}\left[(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}) \leq \epsilon\right] = \Phi(\epsilon)$$

This implies that

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq n\mu + \sqrt{n}\sigma\epsilon\right) = \Phi(\epsilon),$$

and hence

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq \mu + \frac{\sigma\epsilon}{\sqrt{n}}\right) = \Phi(\epsilon).$$

As $n \to \infty$, $\frac{\sigma\epsilon}{\sqrt{n}} \to 0$. Therefore, the deviation that central limit theorem can handle is small deviation.

TO DO: Add a picture to explain small deviation VS large deviation.