Law of Large Numbers and Central Limit Theorem

1 Convergence in Probability and Law of Large Numbers

Definition 1 *A sequence of r.v.s* X_n , $n = 1, 2, 3, \ldots$ *is said to converge to* Y *if and only if for every* $\epsilon > 0$ *then*

$$P(|X_n - Y| > \epsilon) \to 0 \ .$$

In this course we usually are only interested in the limit being a constant, say a. We then say (the sequence) X_n converges to a in probability if and only if

$$P(|X_n - a| > \epsilon) \to 0 \ .$$

(also write $X_n \to a$).

In our Law of Large Numbers application we have the following setting : Y_i are iid with mean μ and variance σ^2. We then consider r.v.s

$$X_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}_n \ .$$

Law of Large Numbers : Suppose X_i are iid r.v.s with finite mean $\mathrm{E}(X_i) = \mu$ and finite variance $\mathrm{Var}(X_i) = \sigma^2$. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \text{ in propbability as } n \to \infty \ .$$

2 Law of Large Numbers: Some Applications

Example Estimate of π.

Suppose that U_1, U_2 are iid Unif(-1,1). This pair of random variables is uniformly distributed over the square $[-1, 1] \times [-1, 1]$. Consider the circle of radius 1, centred at (0,0). Define the random variable

$$X = \begin{cases} 1 & \text{if } U_1^2 + U_2^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $X \sim$ Bernoulli(p) with $p = \frac{\pi}{4}$. Using a pseudo random number generator, we could simulate n pairs of Uniform(-1,1) random variables, $(U_{i,1}, U_{i,2})$ and then obtain X_i as above. Then X_i, $i = 1, \ldots, n$ are iid Bernoulli($p = \frac{\pi}{4}$) r.v.'s. Since X has a finite variance, then by the Law of Large Numbers (LLN)

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \frac{\pi}{4} \ .$$

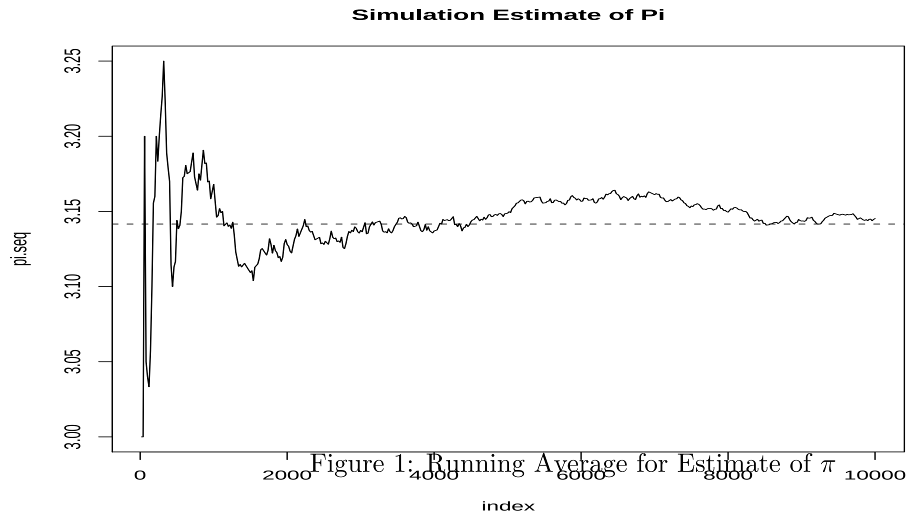

Figure 1: Running Average for Estimate of π

Thus we could use $4\bar{X}_n$ as an estimate of π. This is shown in Figure 1 for $n = 10000$.

The example above is interesting as it relates to other applications in finance and actuarial science.

The underlying random variable is a two dimensional random vector (U_1, U_2). The LLN is not applied directly to this but to a function of this random vector, something of the form $Y = h(U_1, U_2)$. Since the random vectors $(U_{i,1}, U_{i,2})$ are iid random vectors then $Y_i = h(U_{i,1}, U_{i,2})$ are iid; the student should recall or review our material study of independence of r.v.s and in particular the theorem about functions of collections of independent r.v.s.

This same idea applies to topics such as the Monte Carlo method to approximate the expected payoff of a stock option. The stock process to model stock prices from time 0 to T is a random vector $(S_0, \ldots, S_T)$. The payoff for an option is a random variable $Y = h(S_0, \ldots, S_T)$. In the special case of a European option this function has only one argument and the payoff r.v. is $Y = h(S_T)$.

If we have a stochastic (statistical) model to simulate an m-th sample path $(S_{m,0}, \ldots, S_{m,T})$, and the simulated paths are independent, then $Y_m = h(S_{m,0}, \ldots, S_{m,T})$ are iid. If we need to condition on an initial value then we require that the sample paths are conditionally independent given these initial conditions, then the Y_m are conditionally independent given the initial conditions. As long the LLN applies then

$$\bar{Y}_M = \frac{1}{M}\sum_{m=1}^{M} Y_m = \frac{1}{M}\sum_{m=1}^{M} h(S_{m,0}, \ldots, S_{m,T}) \to E_P\left(h(S_0, \ldots, S_T) \mid \text{initial conditions}\right)$$

as $M \to \infty$. In this last expression the subscript P on the expectation sign is to indicate that this expectation is with respect to the model that we are using for the simulation. In terms of the mathematics of option pricing this may be the so called *historical measure* or it may be the so called *risk neutral measure*.

In an actuarial example we would wish to model and simulate the random process to produce the r.v. that is the payoff for a particular policy.

Example Cauchy Example

What happens when there is no first moment. Such an example is the Cauchy distribution. It can be shown (but beyond this course) that if X_i, $i \geq 1$ are iid Cauchy, then $\bar{X}_n$ does not converge in distribution to any constant. In fact $\bar{X}_n \sim$ Cauchy. This is illustrated in Figure 2. We see this plot gives not indication of $\bar{X}$ settling in on any particular value. The plot every once in a while takes big jumps.

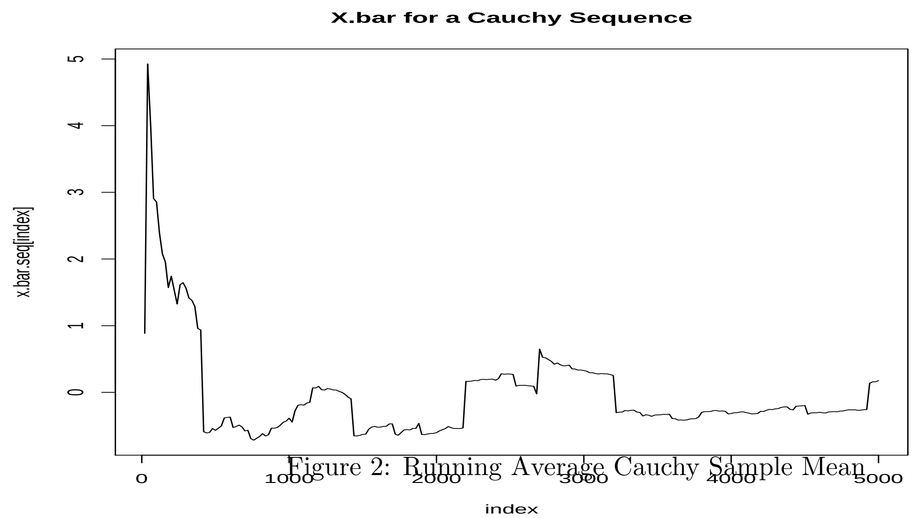

Figure 2: Running Average Cauchy Sample Mean

The fact that $\bar{X}_n$ has a Cauchy distribution can be studied by the convolution formula. However to evaluate, that is obtain a formula for this integral even in the case $n = 2$ requires complex integration and the residue theorem from complex variables.

Example Monte Carlo Integration

Suppose we wish to evaluate an integral $\int_a^b f(x)dx$. There are various numerical ways of doing this. However one particularly useful method is Monte Carlo integration. Suppose that $X \sim \text{Unif}(a, b)$. Then

$$\mathrm{E}(f(X)) = \frac{1}{b-a}\int_a^b f(x)dx\ .$$

Thus if we generate X_i, $i = 1, \ldots, n$ iid Uniform(a, b) r.v.'s, then by the Law of Large Numbers

$$(b-a)\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \int_a^b f(x)dx\ .$$

Consider $f(x) = \sin(x)$, and evaluate the integral $\int_0^4 f(x)dx$. This particular example can be evaluated analytically giving

$$\int_0^4 f(x)dx = 1 - \cos(4) = 1.65364$$

Here we use $a = 0$, $b = 4$. This is illustrated in Figure 3.

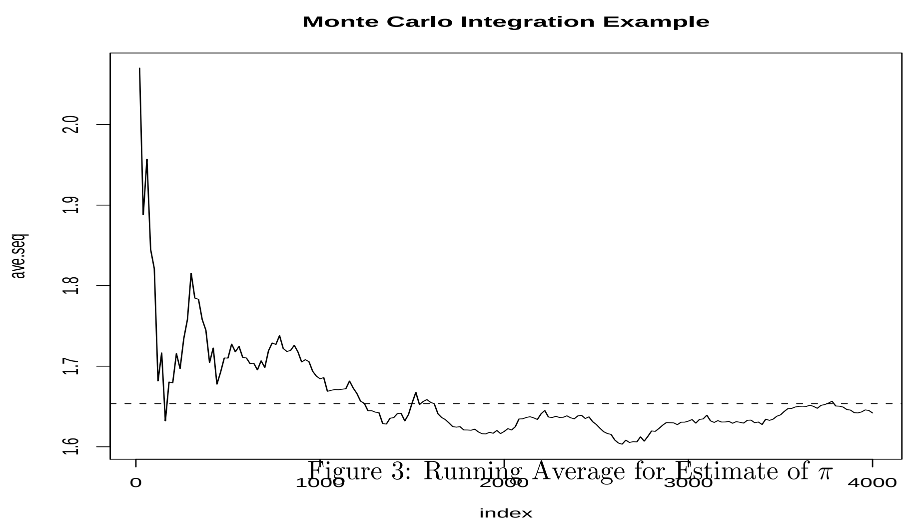

Figure 3: Running Average for Estimate of π

Example Empirical Distribution Function (EDF)

Suppose that X_i, $i = 1, 2, 3, \ldots$ are i.i.d. with cdf F. Consider the (random) function F_n which maps $R \mapsto [0, 1]$ and is given by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{(-\infty,x]}(X_i)$$

Notice that F_n obeys all the properties of being a cdf. It is called the empirical distribution function of the data $X_1, X_2, \ldots, X_n$.

Notice that, for a fixed x, $Y_i = \mathrm{I}_{(-\infty,x]}(X_i)$, $i = 1, 2, 3, \ldots$ are i.i.d. Bernoulli(p) with $p = F(x)$. For given x, y the random variables $Y_i = \mathrm{I}_{(-\infty,x]}(X_i)$ and $W_i = \mathrm{I}_{(-\infty,y]}(X_i)$ are dependent r.v.'s. Thus (Y_i, W_i) are i.i.d. bivariate random variables.

The LLN applies to the sequence of r.v.'s $Y_i = \mathrm{I}_{(-\infty,x]}(X_i)$ and thus we have

$$F_n(x) \to F(x) \text{ in probability as } n \to \infty$$

This result is fundamental to the reason that statistical inference works. With an infinite amount of iid data we can then calculate $F(x)$ exactly for an argument x. Another aspect of statistical inference deals with the question of how can one estimate distributions or parameters with finite amounts of data from iid experiments.

3 Central Limit Theorem: Some Applications

The Central Limit Theorem (CLT) applies to i.i.d. sequences with finite mean and variance.

Theorem 1 (Central Limit Theorem) *Suppose that $X_i, i \geq 1$ is an i.i.d sequence of r.v.'s with finite mean μ and finite variance $\sigma^2 > 0$. Let*

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}$$

be the standardized variable obtained from

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

Then $Z_n \Rightarrow N(0,1)$ (Z_n converges in distribution to standard normal) as $n \to \infty$.

For a given $q \in (0,1)$ let $z_q = \Phi^{-1}(1-q)$ be the $1-q$ quantile of the standard normal distribution. If the CLT applies then

$$\begin{aligned} 1-q &= P(Z_n \leq x) \\ &\to \Phi(x) \end{aligned}$$

and so x satisfies (approximately) $1-q = \Phi(x)$. Hence $x = z_q$ to the accuracy of this approximation. Thus we can use the CLT to obtain approximate quantiles for the distribution of Z_n. With some straightforward manipulations we can then also obtain approximate quantiles for the distribution of $\bar{X}_n$ or of $\sum_{i=1}^{n} X_i$.

Application : Asymptotic (or Approximate) Prediction Intervals and Confidence Intervals

First let us consider the exact confidence interval for a population mean. In this part we are only dealing with distributions with finite means and variances. Specifically consider $X_i, i = 1, \ldots, n$ iid from from a distribution, say F. Consider the r.v. $\bar{X}_n$. The 95% confidence interval for $\mu = \mathrm{E}(X_1)$ is given by

$$\left\{\mu \mid a \leq \bar{X}_n - \mu \leq b\right\}$$

where a, b are the 0.025 and 0.975 quantiles of the distribution of $\bar{X}_n$. How can we calculate these values? For this we need to obtain the distribution of $\bar{X}_n$, which requires an $n-1$ dimension integral (or sum) or $n-1$ iterations of convolution formulae. However we can approximate these quantiles as follows.

$$\{\mu \mid a \leq \bar{x} - \mu \leq b\} \quad = \quad \left\{\mu \mid \frac{\sqrt{n}a}{\sigma} \leq \frac{\sqrt{n}\left(\bar{x} - \mu\right)}{\sigma} \leq \frac{\sqrt{n}b}{\sigma}\right\}$$

By the CLT

$$P\left(\frac{\sqrt{n}\left(\bar{x} - \mu\right)}{\sigma} \leq z\right) \to \Phi(z) \equiv P(Z \leq z)$$

as $n \to \infty$. Since solving $P(Z \leq z) = 0.975$ gives solution $z = 1.96$ we thus have approximately

$$\frac{\sqrt{n}b}{\sigma} = 1.96 \ .$$

Similarly we can approximate a.

Below we consider this a bit more generally.

As an example, consider for a given $\alpha \in (0, 1/2)$, solve for x in $P(|Z_n| > x) = \alpha$. Then by the CLT $x = z_{\alpha/2}$. For example if $\alpha = .05$ we obtain $x = z_{.025} = 1.96$. Thus with probability approximately $1 - \alpha$

$$\begin{aligned} -z_{\alpha/2} \leq Z_n \leq z_{\alpha/2} \quad &\Leftrightarrow \quad -z_{\alpha/2} \leq \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq z_{\alpha/2} \\ &\Leftrightarrow \quad -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \end{aligned}$$

From this we can then construct

1. central $1 - \alpha$ probability intervals for $\bar{X}_n$
2. (central) $100(1 - \alpha)\%$ confidence intervals for μ (provided σ is known.

The probability $1 - \alpha$ prediction interval for $\bar{X}_n$ is

$$\left[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

This means that $\bar{X}_n$ is in this interval with probability $1 - \alpha$.

The $100(1 - \alpha)\%$ confidence interval for μ is the set of μ which belongs to the set

$$\left\{\mu : -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X}_n - \mu \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

Rewriting this set we obtain

$$\left\{\mu : \bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

or in equivalent form

$$\left[\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \ . \tag{1}$$

The observed value of this confidence interval is then the interval with the observed value of $\bar{X}_n$, say $\bar{X}_{n,\text{obs}}$, obtained from our data substituted in place of $\bar{X}_n$.

There is another theorem related to the CLT. The proof is beyond what we can study in this course.

Theorem 2 *Suppose that $X_i, i \geq 1$ is an i.i.d sequence of r.v.'s with finite mean μ and finite variance $\sigma^2 > 0$. Let*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$$

be the sample mean and sample variance respectively. Let

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n}$$

Then $Z_n \Rightarrow N(0, 1)$ (Z_n converges in distribution to standard normal) as $n \to \infty$.

Notice this Theorem looks exactly like the CLT except that the sample variance is substituted in place of the population variance.

Using Theorem 2 we can then construct a (central) $100(1-\alpha)\%$ confidence interval for μ. Notice that we can do this based on the observed data since $\bar{X}_n$ and s^2 are values we can calculate after a sample of data is observed. In this case we have as our asymptotic $100(1-\alpha)\%$ confidence interval for μ

$$\left\{\mu : \bar{X}_n - z_{\alpha/2}\frac{\sqrt{s_n^2}}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\alpha/2}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right\}$$

or in equivalent form

$$\left[\bar{X}_n - z_{\alpha/2}\frac{\sqrt{s_n^2}}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sqrt{s_n^2}}{\sqrt{n}}\right] \ . \tag{2}$$

The confidence intervals (1) and (2) are random intervals. These are studied in mathematical statistics, under the topic of statistical inference.

Remark : Why would we use an approximation if an exact (no approximation) can be used? Of course we not. There is one very important special case where there is an exact result. It applies in the case where we observe iid random variables $X_i, i = 1, \ldots, n$ from a normal population, that is $X_i \sim N(\mu, \sigma^2)$. Surprisingly it turns out that these results will hold no matter what the specific values of the parameters μ, σ^2 happen to be.

Notice this confidence interval formula is obtained from an asymptotic (that is when $n \to \infty$) approximation to the sampling distribution of Z_n. If additional knowledge is available sometimes we a known or exact form for this distribution and can use that. For example if X_i, $i = 1, \ldots, n$ are iid $N(\mu, \sigma^2)$ then

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n} \sim t_{(n-1)} \ .$$

This may then be used to obtain an *exact* confidence interval. The main difference is that different critical values are used. However the coverage probability is now exactly $1 - \alpha$.