ORF 245 Fundamentals of Statistics
Chapter 5
Limit Theorems

Robert Vanderbei

Fall 2014

Slides last edited on December 2, 2014
Law of Large Numbers

Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent identically distributed random variables with mean $\mu$ and variance $\sigma^2$. Let

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

Compute the mean and variance of $\bar{X}_n$:

$$\mathbb{E}(\bar{X}_n) = \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu$$

$$\text{Var}(\bar{X}_n) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \sigma^2 / n$$

The Law of Large Numbers says that

$$\lim_{n \to \infty} \bar{X}_n = \mu$$

Technically speaking there are two versions of this theorem: the strong version and the weak version. Such a distinction is important in more advanced classes. We'll ignore it.
Suppose we want to integrate some function over some interval. For example, suppose that the function is \( f(x) = 1/x \) and the interval is \([1, 4] \):

\[
I = \int_1^4 \frac{1}{x} dx
\]

There are various methods. First choice would be to do an explicit calculation:

\[
I = \int_1^4 \frac{1}{x} dx = \ln(4) - \ln(1) = \ln(4)
\]

But, is this really “explicit”? What is the numerical value of \( \ln(4) \)? Saying there is a button for it on your calculator is not an acceptable answer. Someone at the company that made your calculator had to code up an algorithm to do the computation.

In your calculus class, you learned how to approximate integrals using various methods of discretization: the Midpoint Rule, the Trapezoidal Rule, Simpsons Rule, etc. Your calculator uses one of these methods.
Monte Carlo Integration – Continued

Here’s another method: generate independent uniformly distributed random variables, $X_1, X_2, \ldots, X_n$ on $[1, 4]$ and compute

$$I_n = \frac{1}{n} \sum_{i=1}^{n} \frac{3}{X_i}$$

The answer is a random variable. But, as $n$ tends to infinity, this random variable converges to the mean value of $3/X$:

$$I_n \longrightarrow \mathbb{E} \left( \frac{3}{X} \right) = \int_{1}^{4} \frac{3}{x} \frac{1}{3} dx = \int_{1}^{4} \frac{1}{x} dx$$

% Monte Carlo
n = 300;
X = random('unif', 1, 4, [1 n]);
I = (4-1)*sum(1./X)/n;
sprintf('%10.7f', I)

1.4554425

% Numerical Integration
x = ((101:400)-0.5)/100;
dx = 1/100;
I = sum(1./x)*dx;
sprintf('%10.7f', I)

1.3862905

% Matlab's Answer
I = log(4);
sprintf('%10.7f', I)

1.3862944

With $n = 3,000,000$, Monte Carlo gives $1.3862486$ (five sig. figs.).

Conclusion: Monte Carlo is a method of last resort.
Central Limit Theorem

Let \( X_1, X_2, \ldots \) be a sequence of independent identically distributed (iid) random variables having mean \( \mu \) and variance \( \sigma^2 \). Let

\[
S_n = \sum_{i=1}^{n} X_i
\]

We know that the expected value of a sum is the sum of the expected values and that the variance of a sum of independent r.v.s' is the sum of the variances. Hence,

\[
\mathbb{E}(S_n) = n\mu \quad \text{and} \quad \text{Var}(S_n) = n\sigma^2
\]

Therefore, the random variable

\[
T_n = \frac{S_n - n\mu}{\sqrt{n} \sigma}
\]

has mean zero and variance one.

The Central Limit Theorem says that, as \( n \) goes to infinity

\[
P(a \leq T_n \leq b) \to \int_{a}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx
\]

In other words, for \( n \) large, \( T_n \) is approximately a \( N(0, 1) \) random variable.
Poisson Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent Poisson random variables with parameter $\lambda = 1$ and let

$$S_n = \sum_{i=1}^{n} X_i$$

We saw before that the sum of independent Poisson random variables is again Poisson. Hence, $S_n$ is Poisson with parameter $\lambda = n$. Hence, it has mean $n$ and variance $n$.

By the Central Limit Theorem, the distribution should look like a Normal with the same mean and variance when $n$ gets large.

```matlab
x = [(0:(2*n))-0.01 0:(2*n)];
x = sort(x);
P = cdf('poiss', x, n);
N = cdf('norm', x, n, sqrt(n));
plot(x,N,'k-'); hold on;
plot(x,P,'b-'); hold off;
```
Binomial Random Variables

Let $X_1, X_2, \ldots$ be a sequence of independent Bernoulli rv’s with parameter $p = 0.4$ and let

$$S_n = \sum_{i=1}^{n} X_i$$

The sum of $n$ independent Bernoulli random variables is Binomial with parameters $n$ and $p$. Hence, it has mean $np$ and variance $npq$, where $q = 1 - p$.

By the Central Limit Theorem, the distribution should look like a Normal with the same mean and variance when $n$ gets large.

```matlab
x = [-2:0.01:n];
N = pdf('norm', x, n*p, sqrt(n*p*q));
plot(x,N,'k-'); hold on;
x = [-2:n];
B = pdf('bino', x, n, p);
bar(x,B,0.1,'FaceColor','b','EdgeColor','b'); hold off;
```
Let \( X_1, X_2, \ldots \) be a sequence of independent exponential random variables with parameter \( \lambda \) and let

\[
S_n = \sum_{i=1}^{n} X_i
\]

We saw before that the sum of \( n \) independent exponential random variables is a Gamma random variable with parameters \( n \) and \( \lambda \). Hence, it has mean \( n/\lambda \) and variance \( n/\lambda^2 \).

By the Central Limit Theorem, the distribution should look like a Normal with the same mean and variance when \( n \) gets large.

\[
x = [-1:0.01:1+(2*n/lambda)];
G = pdf('gam', x, n, 1/lambda);
N = pdf('norm', x, n/lambda, sqrt(n)/lambda);
plot(x,G,'b-'); hold on;
plot(x,N,'k-'); hold off;
\]

\[
x = [-1:0.01:1+(2*n/lambda)];
G = pdf('gam', x, n, 1/lambda);
N = pdf('norm', x, n/lambda, sqrt(n)/lambda);
plot(x,G,'b-'); hold on;
plot(x,N,'k-'); hold off;
\]
Recall: Standard and Poors 500 – Daily Returns

\[ \mu = \mathbb{E}(R_i) \approx \sum_j R_j/n = 9.86 \times 10^{-4}, \quad \sigma^2 = \text{Var}(R_i) \approx \sum_j (R_j - \mu)^2/n = 0.0108 \]

Why the excellent match? *The CENTRAL LIMIT THEOREM!*