

© Copyright 2004-2019 by Jenny A. Baglivo. All Rights Reserved.

| | | |
|----------|---|----------|
| 5 | MATH442601 2 Notebook 5 | 3 |
| 5.1 | Sequences of IID Random Variables | 3 |
| 5.1.1 | Running Sums and Running Averages | 3 |
| 5.2 | Law of Large Numbers (LLN) | 4 |
| 5.3 | Central Limit Theorem (CLT) | 7 |
| 5.3.1 | Special Cases | 12 |
| 5.4 | Moment Generating Functions and the Central Limit Theorem | 15 |

5 MATH442601|2 Notebook 5

This notebook is concerned with patterns that emerge as the number of repetitions of an experiment grows large. We will study two important theorems: the law of large numbers, and the central limit theorem. The notes correspond to material in Chapter 5 of the Rice textbook.

5.1 Sequences of IID Random Variables

Let X_1, X_2, X_3, \dots be a sequence of mutually independent, identically distributed (IID) random variables, each with the same distribution as X .

5.1.1 Running Sums and Running Averages

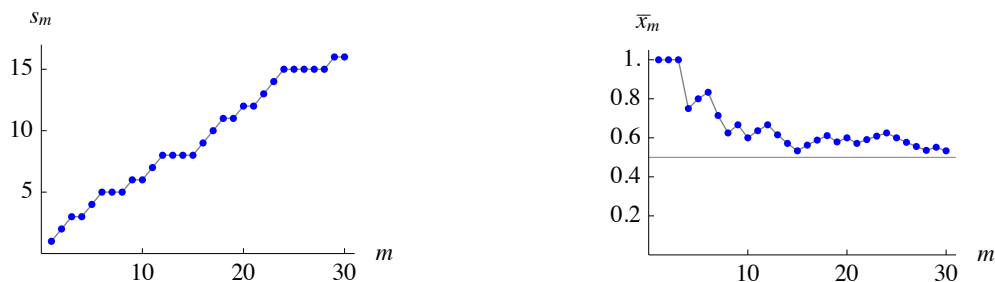
Two related sequences are of interest:

1. Sequence of Running Sums: $S_m = \sum_{i=1}^m X_i$, for $m = 1, 2, 3, \dots$
2. Sequence of Running Averages: $\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i$, for $m = 1, 2, 3, \dots$

Example. For example, consider tossing a fair coin and recording $X = 1$ if a head appears and $X = 0$ otherwise. I used the computer to simulate 30 coin tosses, and here are the results:

1, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0

The observed sequences of running sums (*left*) and averages (*right*) are shown below:



Note that X is a Bernoulli random variable with $p = \frac{1}{2}$ and, for each m , S_m is a binomial random variable based on m trials of a Bernoulli experiment with success probability $\frac{1}{2}$.

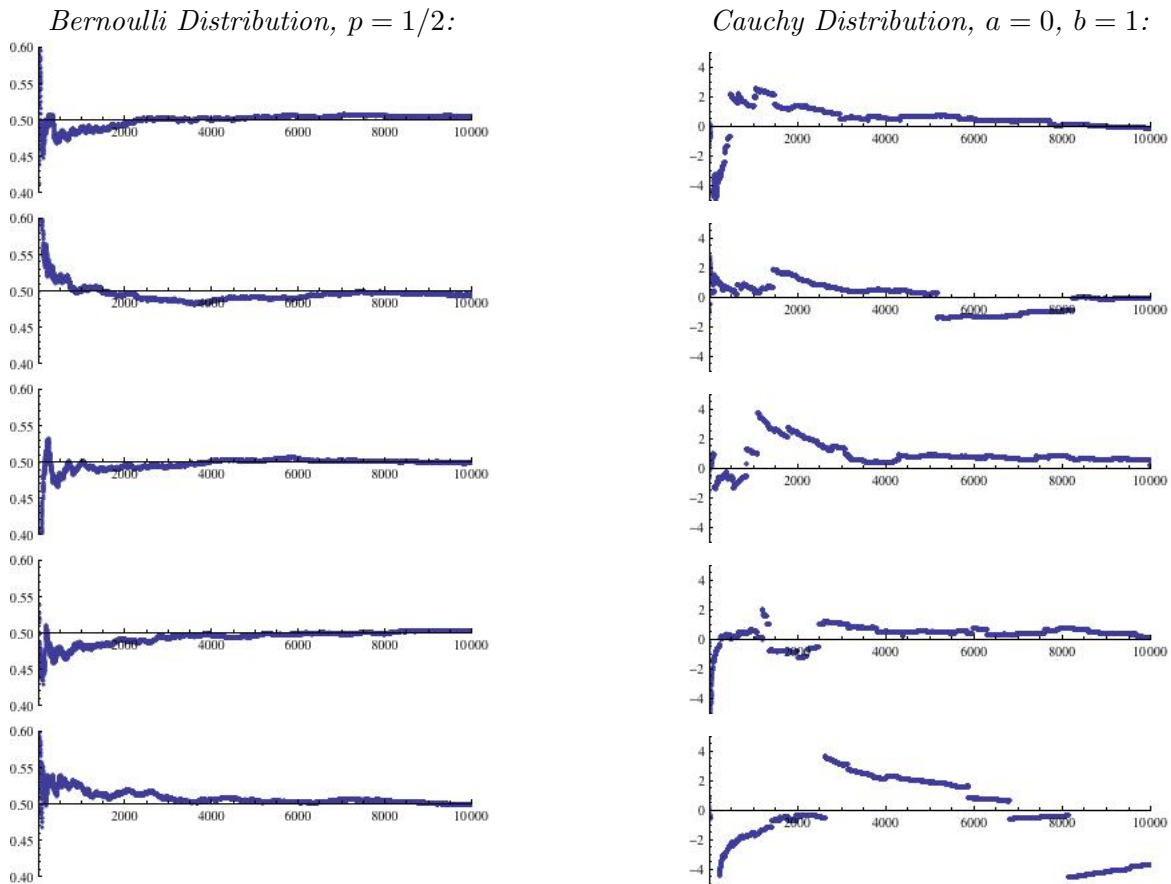
5.2 Law of Large Numbers (LLN)

For distributions with finite mean and standard deviation, the law of large numbers says that the sample mean, \bar{X} , is unlikely to be far from the true mean, μ , when the sample size is large enough. Formally,

Theorem (Law of Large Numbers). Let X be a random variable with mean $\mu = E(X)$ and standard deviation $\sigma = SD(X)$, and let ϵ be a positive real number. Then

$$\lim_{m \rightarrow \infty} P(|\bar{X}_m - \mu| \geq \epsilon) = 0.$$

Example. To illustrate the law of large numbers, I used the computer to simulate 10000 repetitions of a Bernoulli experiment and of a Cauchy experiment. The *left* column of plots is five observed sequences of running averages from the Bernoulli experiment, and the *right* column is five observed sequences of running averages from the Cauchy experiment.



Left Plots (Bernoulli): The running averages are almost identically equal to $\frac{1}{2}$ for large m .

Right Plots (Cauchy): Running averages can be quite far from the center of the Cauchy distribution, however, since this distribution has no mean and no standard deviation.

Exercise. One form of the Chebyshev inequality says that

$$P\left(|Y - \mu_y| \geq k\sigma_y\right) \leq \frac{1}{k^2},$$

where Y is a random variable with mean $\mu_y = E(Y)$ and standard deviation $\sigma_y = SD(Y)$, and k is a positive constant. Use this form to prove the law of large numbers.

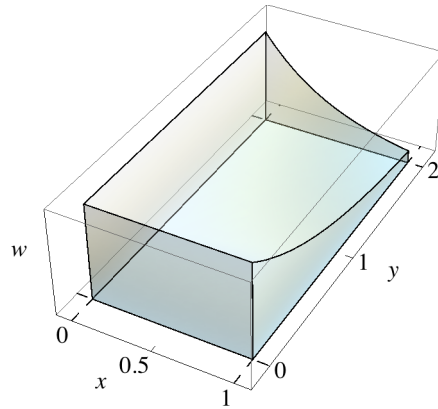
Monte Carlo Evaluation of Integrals. An interesting application of the law of large numbers is to the approximation of multiple integrals.

Example. For example, consider evaluating the double integral

$$\int_0^1 \int_0^2 e^{-xy} dy dx,$$

and let $\mathcal{R} = [0, 1] \times [0, 2]$ be the region of integration.

[Note that the value of the double integral is the volume of the solid shown on the right.]



Assume that (X, Y) has a bivariate uniform distribution on \mathcal{R} , and let $W = e^{-XY}$.

1. Since the area of \mathcal{R} is 2, the expected value of W is

$$\mu = E(W) = E(e^{-XY}) = \int_0^1 \int_0^2 e^{-xy} \left(\frac{1}{2}\right) dy dx = \frac{1}{2} \left(\int_0^1 \int_0^2 e^{-xy} dy dx \right),$$

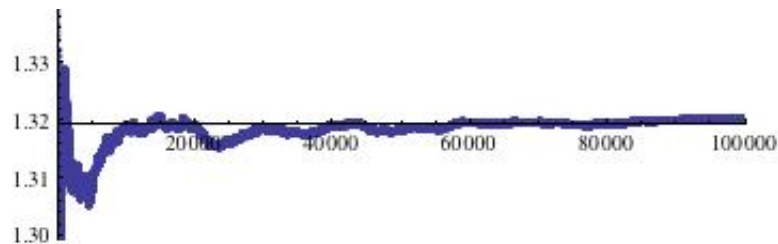
and 2μ is the double integral we want.

2. To approximate the double integral using Monte Carlo simulation, consider the sequence

$$2\bar{W}_m = \frac{2}{m} \sum_{i=1}^m e^{-X_i Y_i}, \quad \text{for } m = 1, 2, 3, \dots,$$

where the (X_i, Y_i) are drawn from the bivariate uniform distribution on the rectangle.

3. I used the computer to simulate 100,000 repetitions of the bivariate uniform experiment, and I plotted the sequence of observed estimates $2\bar{w}_m$, for $m = 1, 2, 3, \dots, 100000$.



5.3 Central Limit Theorem (CLT)

The most important theorem in a probability course is the central limit theorem. Its proof is attributed to P. Laplace and A. deMoivre.

Theorem (Central Limit Theorem). Let X_1, X_2, X_3, \dots be a sequence of independent, identically distributed random variables, each with the same distribution as X . Assume that the moment generating function of X converges in an open interval containing zero, and let $\mu = E(X)$ and $\sigma = SD(X)$. Then, for every real number x ,

$$\lim_{m \rightarrow \infty} P\left(\frac{S_m - m\mu}{\sigma\sqrt{m}} \leq x\right) = \Phi(x),$$

where $S_m = \sum_{i=1}^m X_i$, and $\Phi()$ is the CDF of the standard normal random variable.

Notes:

1. Convergence in Distribution: The result of the central limit theorem is an example of convergence in distribution. Specifically, let $Z_m = \frac{S_m - m\mu}{\sigma\sqrt{m}}$, for $m = 1, 2, 3, \dots$, be the sequence of standardized running sums, and let $F_m(x) = P(Z_m \leq x)$ be the CDF of Z_m , for each m . The central limit theorem says that the sequence of CDF's converges to the CDF of the standard normal random variable. That is,

$$F_m(x) \rightarrow \Phi(x), \text{ as } m \rightarrow \infty.$$

Thus, we say that Z_m converges in distribution to the standard normal random variable.

2. Continuity Theorem: The proof of the central limit theorem uses a property of moment generating functions known as the continuity theorem.

Continuity Theorem. Let $\{F_m(x)\}$ be a sequence of cumulative distribution functions for random variables with corresponding moment generating functions $\{M_m(t)\}$, and let $F(x)$ be the CDF of a random variable whose moment generating function is $M(t)$. If

$$M_m(t) \rightarrow M(t) \text{ for all } t \text{ in an open interval containing } 0,$$

then $F_m(x) \rightarrow F(x)$ at all continuity points of the CDF $F(x)$.

In our application of the continuity theorem, $M_m(t)$ is the moment generating function of the standardized sum $Z_m = \frac{S_m - m\mu}{\sigma\sqrt{m}}$, and we can demonstrate that

$$M_m(t) \rightarrow e^{t^2/2} \text{ for all } t \text{ in an open interval containing } 0.$$

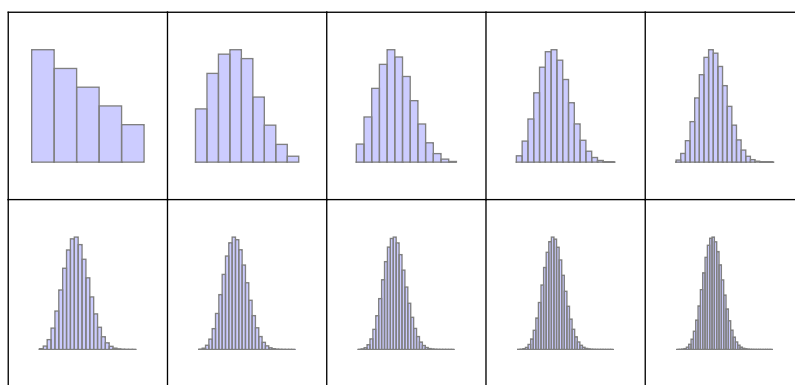
Since $M(t) = e^{t^2/2}$ is the moment generating function of the standard normal random variable, the result follows.

3. Normal Approximation to a Sample Sum: Since Z_m converges in distribution to the standard normal, we can use probabilities based on the standard normal random variable to approximate probabilities for sums of independent, identically distributed (IID) random variables when m is large enough. That is,

$$P(S_m \leq s) \approx \Phi\left(\frac{s - m\mu}{\sigma\sqrt{m}}\right), \text{ for each } s, \text{ when } m \text{ is large enough.}$$

4. Rate of Convergence: The central limit theorem tells us nothing about the rate of convergence of the sequence of CDFs. The graph below illustrates the speed of convergence for the finite discrete random variable X whose PMF is as follows:

| | | | | | |
|--------|------|------|------|------|------|
| x | 1 | 2 | 3 | 4 | 5 |
| $p(x)$ | 0.30 | 0.25 | 0.20 | 0.15 | 0.10 |



The plot shows the probability histograms of $S_1 = X, S_2, S_3, \dots, S_{10}$. Note that the distribution of S_{10} is approximately bell-shaped.

The convergence for other types of distributions may take much longer. For example, if the distribution of X is very skewed, then the convergence will be slow. Thus, we need to be cautious when using normal approximations.

5. Normal Approximation to a Sample Mean: If we divide the numerator and denominator of the standardized sum Z_m by m , we see that

$$Z_m = \frac{S_m - m\mu}{\sigma\sqrt{m}} = \frac{(S_m - m\mu)/m}{(\sigma\sqrt{m})/m} = \frac{\bar{X}_m - \mu}{\sigma/\sqrt{m}}, \text{ for } m = 1, 2, 3, \dots$$

Thus, the central limit theorem tells us that we can use probabilities based on the standard normal random variable to approximate probabilities for averages of independent, identically distributed (IID) random variables when m is large enough. That is,

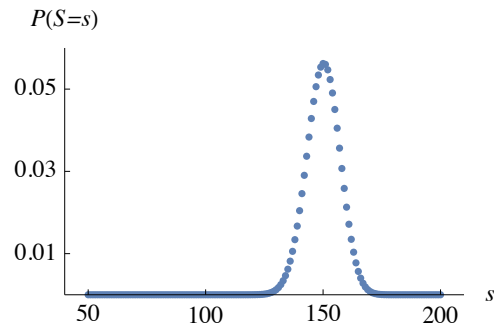
$$P(\bar{X}_m \leq \bar{x}) \approx \Phi\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{m}}\right), \text{ for each } \bar{x}.$$

6. Normal Random Variable X : If X is a normal random variable, the formulas in items 3 and 5 are exact. That is, the distributions of the sample sum and of the sample mean are exactly normal for all m .

Exercise. Let X be the discrete random variable with PMF

| | | | | |
|--------|------|------|------|------|
| x | 1 | 2 | 3 | 4 |
| $p(x)$ | 0.10 | 0.20 | 0.30 | 0.40 |

where $p(x) = 0$ otherwise, and let S be the sample sum of a random sample of size 50 from the X distribution. [Note: A dot plot of pairs $(s, P(S = s))$ is shown to the right.]



Use the central limit theorem to approximate $P(120 \leq S \leq 160)$.

Exercise. The Quick Delivery Package Service can hold up to 120 packages in its new Puddle Jumper aircraft. Let X equal the weight (in pounds) of a package and assume that X has a continuous distribution with PDF

$$f(x) = \frac{1}{2592} x \text{ when } 0 \leq x \leq 72, \text{ and } 0 \text{ otherwise.}$$

Given a random sample of size 120 from the X distribution, use the appropriate large sample approximation to estimate the probability that

- (a) The total weight is more than 6000 pounds.
- (b) The mean weight is between 45 and 51 pounds.

Exercise. Suppose that calls received by a medical hotline during the night shift (midnight to 8AM) follow a Poisson process with an average rate of 5.5 calls per shift. Let \bar{X} be the sample mean number of calls received by per night shift over a 28-day period. Use the appropriate large sample approximation to estimate the probability that \bar{X} is between 4.5 and 6.1.

5.3.1 Special Cases

The central limit theorem implies that the distributions of certain binomial, negative binomial, Poisson and gamma random variables can be approximated by normal distributions.

Binomial Distribution: Let X be a binomial random variable with parameters n and p . Since X can be thought of as the sum of n independent Bernoulli random variables, each with success probability p , the distribution of X can be approximated by a normal distribution with mean np and standard deviation $\sqrt{np(1-p)}$ when n is large.



Negative Binomial Distribution: Let X be a negative binomial random variable with parameters r and p . Since X can be thought of as the sum of r independent geometric random variables, each with success probability p , the distribution of X can be approximated by a normal distribution with mean $\frac{r}{p}$ and standard deviation $\sqrt{\frac{r(1-p)}{p^2}}$ when r is large.

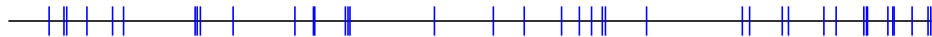


Poisson Distribution: Consider events occurring over time according to a Poisson process. Let X be the number of events occurring in a fixed time period, and assume that we expect λ events for this time period. Then X is a Poisson random variable with parameter λ .

If we subdivide the interval into, say, m subintervals, then X can be thought of as the sum of m independent Poisson random variables, each with parameter $\frac{\lambda}{m}$. Thus, X can be approximated by a normal distribution with mean λ and standard deviation $\sqrt{\lambda}$ when λ is large.



Gamma Distribution: Let X be the time measured from time 0 to the occurrence of the r^{th} event of a Poisson process whose rate is λ events per unit time. Then X can be thought of as the sum of r independent exponential random variables representing (1) the time to the first event, (2) the time from the first to the second event, (3) the time from the second to the third event, and so forth. Thus, X can be approximated by a normal distribution with mean $\frac{r}{\lambda}$ and standard deviation $\sqrt{\frac{r}{\lambda^2}} = \frac{\sqrt{r}}{\lambda}$ when r is large.



Exercise (The House Special). A roulette wheel has 38 pockets labelled

0, 00, 1, 2, 3, 4, \dots , 36.

Assume that the ball lands in each pocket with equal probability and the results are independent from game to game. The “house special” is a bet on the 5 pockets 0, 00, 1, 2, 3. The bet pays 6-to-1. (If you place a one dollar bet and win, then you get back your dollar plus six more dollars; if you lose, then you get back nothing.)

Assume you make 175 one-dollar bets on the house special. Use the appropriate large sample approximation to estimate the probability that you come out ahead.

Exercise. Assume that the occurrences of small earth tremors in southern California follow a Poisson process, with an average rate of 1 event every 2 months.

Use the appropriate large sample approximation to estimate the probability that the time (starting from today) until the 95th small tremor is 180 months or less.

5.4 Moment Generating Functions and the Central Limit Theorem

As mentioned in the previous section, the proof of the central limit theorem is an application of the continuity theorem. In this section, we go through the steps leading to the key result, namely, that

$$M_m(t) \rightarrow e^{t^2/2} \text{ for all } t \text{ in an open interval containing } 0.$$

Recall that $\mu = E(X)$ and $\sigma = SD(X)$. Now,

Step 1: Rewrite Standardized Sum as a Sum. For a given m ,

$$Z_m = \frac{S_m - m\mu}{\sigma\sqrt{m}} = \frac{(\sum_{i=1}^m X_i) - m\mu}{\sigma\sqrt{m}} = \frac{\sum_{i=1}^m (X_i - \mu)}{\sigma\sqrt{m}} = \sum_{i=1}^m \frac{X_i - \mu}{\sigma\sqrt{m}} = \sum_{i=1}^m Y_i.$$

Step 2: Find the Mean and Variance of Each Summand. For a given m , let

$$Y = \frac{1}{\sigma\sqrt{m}}(X - \mu).$$

Then,

$$E(Y) = \underline{\hspace{20em}}$$

$$Var(Y) = \underline{\hspace{20em}}$$

Step 3: Approximate the Moment Generating Function of Each Summand. For a given m ,

$$M_Y(t) = \sum_{k=0}^{\infty} \frac{M_Y^{(k)}(0)}{k!} t^k \quad (\text{Maclaurin series expansion})$$

$$= \sum_{k=0}^{\infty} \frac{E(Y^k)}{k!} t^k \quad (\text{since } M_Y^{(k)}(0) = E(Y^k))$$

$$= 1 + E(Y)t + \frac{E(Y^2)}{2}t^2 + \frac{E(Y^3)}{3!}t^3 + \frac{E(Y^4)}{4!}t^4 + \dots$$

$$\approx 1 + E(Y)t + \frac{E(Y^2)}{2}t^2 \quad (\text{quadratic approximation})$$

$$= \underline{\hspace{20em}}$$

Step 4: Approximate the Moment Generating Function of the Standardized Sum. Let $M_m(t)$ be the moment generating function of the standardized sum for a given value of m . Then,

$$M_m(t) = \left(M_Y(t)\right)^m \approx \left(\underline{\hspace{15em}}\right)^m$$

Step 5: Limit of the Approximation. (Please complete this step.)