

9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Homework Assignment 6
October 26, 2016
Due November 7, 2016 at 5:00 PM

For each of the simulation problems, turn in both your MATLAB[®] code and the answers to the questions.

1. Prior to the development of a wide range of simulation algorithms, a common approach to generating standard Gaussian random variables was to simulate six uniform random numbers, sum them and normalize them. Suppose that X has the probability density function

$$f(x) = 1 \quad 0 \leq x \leq 1$$

A. Compute the mean μ_x and the variance σ_x^2 of X .

B. Let $Z = \frac{n^{\frac{1}{2}}(\bar{X} - \mu_x)}{\sigma_x}$ where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and the $X_{i,s}$ are independent random draws from $f(x)$. Take $n = 6$ and write a MATLAB program to simulate Z .

C. Use the algorithm in B to simulate 500 draws of Z .

D. Make a Q-Q plot of the Z_s comparing them to a standard Gaussian distribution. Is this a reasonable way to simulate standard Gaussian random variables?

E. Given the algorithm in C, how could you simulate draws from a Gaussian density $f(y)$ with mean μ_y and variance σ_y^2 ? Do not write an algorithm, simply explain how you would obtain the Y_s from the Z_s .

2. The October 24, 2016 NBC/Wall Street Journal poll of 768 likely voters in New Hampshire reported that 45% preferred Hillary Clinton and 36% preferred Donald Trump in the November 8 presidential election. The poll reports an error of $\pm 3.5\%$.

A. Use the Gaussian approximation to the binomial to compute an approximate 95% confidence interval for the proportion of voters who are likely to vote for Secretary Clinton based on a sample of 768. How does the length of your 95% confidence interval compare to the stated error in the poll? How does your result compare with our approximate rule of our 1 over the square root of the sample size rule?

B. If we interpret the error bounds of 7.0% as a 95% confidence interval how many voters would need to be polled based on an analysis using Chebyshev's inequality?

- C. Which approach, Gaussian approximation or Chebyshev's inequality, do you think the NBC/Wall Street Journal pollsters used to determine the sample size for their survey? Explain.

3. An important application of the Law of Large Numbers is computing integrals and areas. Consider the circle $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{4}$ for $x \in (0,1)$ and $y \in (0,1)$. The circle is completely contained within the unit square whose coordinates are $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$. We can use the Law of Large Numbers to estimate its area by Monte Carlo.

A. Plot the circle.

- B. The area of the circle is $\frac{\pi}{4}$ and can be computed by Monte Carlo by drawing (X_i, Y_i) pairs from the unit square. That is draw X_i from $U(0,1)$ and Y_i from $U(0,1)$. Let Z_i be a Bernoulli random variable which is 1 if $(X_i - \frac{1}{2})^2 + (Y_i - \frac{1}{2})^2 < \frac{1}{4}$ and 0 otherwise. The probability that $(X_i - \frac{1}{2})^2 + (Y_i - \frac{1}{2})^2 < \frac{1}{4}$ is $\frac{\pi}{4}$. Hence by the Law of Large Numbers,

$$\frac{\pi}{4} \approx \hat{p} = \frac{\sum_{i=1}^n Z_i}{n}.$$

for n sufficiently large. Write a Monte Carlo algorithm to compute \hat{p} for $n = 100, 500, 1,000$ draws. How does the accuracy of the approximation improve as n increases? (Note that the error of the approximation as a function of n can be evaluated as $\hat{p} \pm 2 \left[\frac{\hat{p}(1-\hat{p})}{n} \right]^{\frac{1}{2}}$.)

4. The Central Limit Theorem (CLT) states that the sum of independent, identically distributed random variables with finite variances will asymptotically (i.e., as the number of random variables gets large) have a Gaussian distribution. This is one reason why approximate Gaussian methods are so common in statistical analyses. In this problem, we explore some features of the CLT for discrete random variables.

- A. Poisson
- i. Plot the distribution of a Poisson random variable with rate parameter values of 1, 5, 15 and 25.
 - ii. Use Q-Q plots to compare this distribution with normal distributions with the appropriate means and variances.
 - iii. Describe how the quality of the Gaussian approximation to the Poisson varies as a function of the parameter.

[Useful MATLAB functions: **sqrt**, **exp**, **cumsum**]

B. Binomial

- i. Plot binomial distributions with different values of n and p . Try a grid with $n = [15, 25, 50]$ and $p = [0.1, 0.3, 0.5]$.
- iv. Use Q-Q plots to compare these distributions with normal distributions with the appropriate means and variances.
- v. Describe how the quality of the Gaussian approximation to the binomial varies as a function of n and p .

[Useful MATLAB functions: **nchoosek**].

5. Assume that a random variable X has a probability density function

$$f(x|\theta) = \frac{[1+\theta x]}{2}$$

for $-1 \leq x \leq 1$, $0 \leq \theta \leq 1$.

- A. Find $E(X)$ and $Var(X)$.
- B. Assume that you observe a random sample x_1, x_2, \dots, x_n from $f(x|\theta)$. Compute the method-of-moments estimate of θ .
- C. Using the Central Limit Theorem find a formula for an approximate 95% confidence interval for the true value of θ based on the results in B.
- D. Find $F(x|\theta)$ and simulate a random sample of size 100 for $\theta = 0.3$.
- E. Use B and C to estimate θ and to construct an approximate 95% confidence interval for the true value of the parameter using the simulated data from D. Does your interval cover the true value of the parameter?

6. Consider X_1, \dots, X_n are independent observations from the Laplace probability density function

$$f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right)$$

A. Assuming that $\sum_{i=1}^n |x_i| = 20$, plot the likelihood function for σ .

B. Find the maximum likelihood estimate of σ .

C. Compute an approximate 95% confidence interval for σ .

7. Consider X_1, \dots, X_n are independent observations from the Pareto probability density function

$$f(x | x_m, \alpha) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}}$$

for $x > x_m$ where x_m is the minimum value of x and $\alpha > 0$ is the shape parameter.

A. Assume that x_m is known. Find the method-of-moments estimate of α .

B. Find the maximum likelihood estimate of α .

C. Compute an approximate 95% confidence interval for α .

8. (Extra-Credit) A Direct Demonstration of the Central Limit Theorem.

In class we proved the Central Limit Theorem using moment generating functions. This is a 4-step exercise to show directly that when λ is large, the probability mass function of a Poisson random variable can be approximated by the probability density of a Gaussian random variable.

Given $\Pr(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$

A. (Easy) Take logs of both sides and use Stirling's approximation to show that

$$\log \Pr(X = k) = -\lambda + k \log \lambda - \left(k + \frac{1}{2}\right) \log k + k - \frac{1}{2} \log(2\pi)$$

Stirling's approximation is $\log n! \approx \left(n + \frac{1}{2}\right) \log n - n + \frac{1}{2} \log(2\pi)$.

B. (Tedious) Let $z = (k - \lambda) / \lambda^{1/2}$. Show that

$$\log \Pr(x = k) \approx -\frac{1}{2} \log(2\pi\lambda) + z\lambda^{\frac{1}{2}} - (\lambda + z\lambda^{\frac{1}{2}} + \frac{1}{2}) \log(1 + z\lambda^{-\frac{1}{2}})$$

C. (Moderate) Recall that for x small the Taylor series expansion about 0 of $\log(1+x)$ is

$$\log(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} + \dots +$$

Use this fact to show that

$$\log \Pr(x = k) \approx -\frac{1}{2} \log(2\pi\lambda) - \frac{z^2}{2} + \lambda^{-\frac{1}{2}} \left(\frac{z^3}{6} - \frac{z}{2} \right) + \lambda^{-1} \left(-\frac{z^4}{3} + \frac{z^2}{4} \right) - \lambda^{-\frac{3}{2}} \frac{z^3}{6}$$

D. If λ is large conclude that

$$\log \Pr(x = k) \approx -\frac{1}{2} \log(2\pi\lambda) - \frac{z^2}{2}$$

or equivalently

$$\Pr(x = k) \approx (2\pi\lambda)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} z^2\right\} = (2\pi\lambda)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \frac{(k-\lambda)^2}{\lambda}\right\}$$

which is the desired result.

MIT OpenCourseWare
<https://ocw.mit.edu>

9.07 Statistics for Brain and Cognitive Science
Fall 2016

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.