February 26, 2020

Math 318 Assignment 6: Due Wed, March 4 at start of class

I. Problems to be handed in:

1. An item experiences bad events according to a Poisson process of rate 2.5 per hour. After 200 bad events, the item must be discarded. Using the Central Limit Theorem, find approximately the probability that an initially new item:

   (a) is discarded within 72 hours,

   (b) has not been discarded within 96 hours.

2. A shipping company needs to transport 300 containers. The weights (in tonnes) of the containers are independent, and each is uniformly distributed on the interval $[2, 8]$. Approximately what load capacity must a ship have in order to be 99% certain that the total weight of the shipment will not exceed the load capacity?

3. This problem concerns problem #5 on Assignment 5.

   (a) The histogram for the number of earthquakes in 10000 samples of 100 decades looks like a bell curve, and it shows that in all but very few samples the number of earthquakes lies between 70 and 130. To understand why this happens:

      i. Explain why the number Y of earthquakes in 100 decades has the same distribution as $\sum_{i=1}^{100} A_i$ where $A_1, \ldots, A_{100}$ are i.i.d. Poisson(1) random variables. What is the mean μ and standard deviation σ of each A_i?

      ii. Apply the central limit theorem to conclude that Y has approximately the same distribution as $100 + 10Z$ where Z has a standard normal distribution.

      iii. How does (ii) explain the bell curve and that its mass is mainly on the interval $[70, 130]$?

   (b) The plot of $M(i)/i$ does not exactly achieve the limiting value 0.040; there is some fluctuation. To understand the magnitude of this fluctuation:

      i. Let $N_j = 1$ if the j^{th} simulation has exactly 100 earthquakes in 100 decades, and otherwise set $N_j = 0$. Then $M(i) = \sum_{j=1}^{i} N_j$. Use the central limit theorem to conclude that, for large i, the distribution of $M(i)$ is approximately that of $\mu i + \sigma\sqrt{i}Z$ for appropriate values of μ and σ (determine these values), where Z has a standard normal distribution.

      ii. Explain why the observed deviation in the plot, between $10^{-4}M(10^4)$ and the limiting value 0.040, is reasonable.

4. This problem concerns the method of Monte Carlo integration, which is a method for the approximate evaluation of an integral $I = \int_0^1 f(x)dx$.

   (a) Let $U_1, \ldots, U_N$ be i.i.d. uniform random variables on the interval $(0, 1)$, and let

   $$I_N = \frac{1}{N}[f(U_1) + \cdots + f(U_N)].$$

   Suppose that $\int_0^1 f(x)^2 dx < \infty$, and let $\sigma^2 = \text{Var} f(U_1) = \int_0^1 f(x)^2 dx - I^2$. Apply the central limit theorem to show that I_N converges to I as $N \to \infty$, in the sense that

   $$P\left(|I_N - I| \le \frac{\sigma x}{\sqrt{N}}\right) \to P(|Z| \le x),$$

   where Z is a standard normal random variable.

   (b) Assuming that $\sigma \le 1$, how large should N be taken to be 95% confident that I_N is within 0.01 of I?

5. Consider the standard normal probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

It is known that there is no closed form for the antiderivative of this function, i.e., for the c.d.f. Φ of the standard normal. However, the c.d.f. can be approximated accurately, and tables give

$$\int_0^1 f(x)\,dx = \Phi(1) - \Phi(0) \approx 0.3413.$$

(a) To demonstrate the method of Monte Carlo integration, use Python to approximate the integral $\int_0^1 f(x)\,dx$ by generating 40000 i.i.d. uniform random numbers on $[0,1]$ and computing the approximation

$$I_{40000} = \frac{f(U_1) + f(U_2) + \cdots + f(U_{40000})}{40000}$$

from #4. Do this three times and record the results from each run. Submit your code and your output.

(b) Another method for approximating this integral is to recall that the integral $\int_0^1 f(x)\,dx$ represents the area underneath the graph of f from $x = 0$ to $x = 1$. To estimate this area, one could simulate a large number of uniform points in the *square* with corners at $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$; then, find the proportion of points that lie underneath the curve $y = f(x)$. Give a short argument (not a proof, just a motivation) as to why this is a reasonable way to approximate this area.

(c) Perform the approximation in (b) by writing code to simulate 40000 i.i.d. uniform random numbers in the square $[0,1] \times [0,1]$ and determine the proportion of them falling in the region $y \le f(x)$. Do this three times and record the results each time. Submit your code and your output.

II. Recommended problems: These provide additional practice but are not to be handed in.

A. (i) Let $Z_1, \ldots, Z_n$ be i.i.d. standard normal random variables, and let $Y = \sum_{i=1}^n Z_i^2$. (Y is said to have a *chi-squared distribution with n degrees of freedom*; it is used in statistics.) Find the moment generating function of Y. In particular, show that if $n = 2$ then Y has an exponential distribution, and give the parameter. [$(1-2t)^{-n/2}$]

(ii) Let Z be a standard normal random variable. Find all the moments of Z (that is, $E(Z^n)$ for $n = 1, 2, \ldots$). (Hint: expand the moment generating function or the characteristic function as a Taylor series). [$E(Z^{2k}) = \frac{(2k)!}{2^k k!}$, $E(Z^{2k+1}) = 0$.]

B. Chapter 2, 80* (=70* in 10^{th} ed.).

Quote of the week: *I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.*

Francis Galton describing the Central Limit Theorem in *Natural Inheritance* (1889).