Programming Assignment 6: Software measurement and Confidence Intervals

Assigned: Nov. 30
Due: Dec. 14

Construct a MATLAB function $\text{MyFun}(N)$ (what if anything it does is up to you) that constructs a random instance of a problem of some kind of size $N$ and carries out some corresponding computation. The content though not the size of the problem should be set by some random process; for example, “Add up $N$ random numbers” or “Invert a random $N \times N$ matrix” The MATLAB utilities “tic” and “toc” or “cputime” can be used to measure running time.

**Problem 1**

Write a function $\text{AnalyzeRunningTimes}(M,N)$ that carries out the following operations

A. Execute $\text{MyFun}(N)$ $M$ times, and record the running times of each run.

B. Determine the mean $\mu$ and standard deviation $\sigma$ of the running times in (A). Print out these values (this is easily done, of course, just by leaving off the semicolon at the end of the statements where they are computed.)

C. Consider the 7 intervals $[\mu - 3.5\sigma, \mu - 2.5\sigma]$, $[\mu - 2.5\sigma, \mu - 1.5\sigma]$, $[\mu - 1.5\sigma, \mu - 0.5\sigma]$, $[\mu - 0.5\sigma, \mu + 0.5\sigma]$, $[\mu + 0.5\sigma, \mu + 1.5\sigma]$, $[\mu + 1.5\sigma, \mu + 2.5\sigma]$, $[\mu + 2.5\sigma, \mu + 3.5\sigma]$. On a single plot, show (a) a histogram divided into these intervals on the $x$-axis, and the fraction of running times that fall into each interval on the $y$-axis; (b) the product $\sigma$ times the Gaussian with mean $\mu$ and standard deviation $\sigma$; that is, $(1/\sqrt{2\pi})e^{-(x-\mu)^2/2\sigma^2}$. 

**Problem 2**

Write a function $\text{CompareRunningTimes}(M,N_1,N_2)$, which does the following:

- Construct a vector $\vec{V}$ of $M$ running times of $\text{MyFun}(N_1)$ and a vector $\vec{W}$ of $M$ running times of $\text{MyFun}(N_2)$

- Compute and print out the means and standard deviations of $\vec{V}$ and $\vec{W}$: $\mu_V$, $\mu_W$, $\sigma_V$, $\sigma_W$.

- Compute and print out the quantities $\mu_W - \mu_V$, $\sqrt{\sigma_V^2 + \sigma_W^2}$. Note: This is the mean and standard deviation of the random variable $X_2 - X_1$ where $X_1$ and $X_2$ are the running times for $\text{MyFun}(N_1)$ and $\text{MyFun}(N_2)$. If this mean is substantially greater than this standard deviation, then that is evidence that the two distributions are different.

- Using the method of permuting labels described on p. 2-5 of *Statistics is Easy*, compute the confidence that the difference between the means $\mu_V$ and $\mu_W$ is significant.