CSE 4214 :: Lab 1

This lab will introduce you to expected values and random processes.

Generating arbitrary random variables

Let x be a discrete-valued random variable, taking values on 1, 2, …, 6, with probability mass function p(x).

- MATLAB provides a routine, `rand`, which generates uniformly distributed random variables on the range from 0 to 1. Given p(x), propose a way to generate instances of x, with probabilities p(x), from `rand`.

- Write a MATLAB function, called `xrand`, implementing the method you describe. The routine takes a 1 by 6 vector, where the first element of the vector is p(1), the second is p(2), and so on. The routine returns a value on 1, 2, …, 6 at random according to the probabilities p(x).

Analytical versus empirical expectation

Let x be a discrete-valued random variable, with probability mass function p(x). For some function f(x) of x, the expected value of f(x) is written E[f(x)].

We can calculate the expected value in two ways. First, we can calculate E[f(x)] analytically, as we did in class. That is,

$$E[f(x)] = \sum_x f(x)p(x).$$

Second, we can calculate E[f(x)] empirically. That is, we can generate n instances of the random variable x (call them $x_1, x_2, \ldots, x_n$), calculate $f(x_1), f(x_2) \ldots, f(x_n)$, and then take the sample mean:

$$E[f(x)] = \frac{1}{n}\sum_{i=1}^{n} f(x_i).$$

If n is sufficiently large (e.g., thousands), the two methods should agree closely (as long as f(x) is not pathological).

Remember from the lecture that both mean and variance can be represented as expectations: mean as E[x], and variance as $E[x^2] - (E[x])^2$.

- Write two MATLAB functions:
  - `anex` – takes a 1 by 6 vector representing p(x), and returns the analytical mean and variance of the distribution p(x) in a 1 by 2 vector (first element is mean, second element is variance).
  - `emex` – takes a 1 by 6 vector representing p(x) and a number n of instances, and returns the empirical mean and variance for n instances of the random variable x, in a 1 by 2 vector (same as for `anex`).

In both cases, your MATLAB functions should call the function `xrand` that you developed in the previous section.

- Let p(x) = 1/6 for all x. Plot the values (mean and variance) returned by `emex` with respect to n on a semilog (x axis) scale, for n = 10, 40, 100, 400, 1000, 4000, 10000. Include a straight line representing the value returned by `anex` on the same plot.
- Considering your plot, discuss the accuracy of `emex`.

Deliverables

Your deliverables for this lab are as follows:

- All MATLAB code for `xrand`, `anex`, and `emex`.
- A written description of the method for generating x, which you implemented in `xrand`.
- Your plot and discussion of the accuracy of `emex`.

Deliverables are due at the end of the lab period on September 25, 2009.

Demonstration

In the lab, the TA will give you a particular p(x). You will demonstrate the operation of your routines `xrand`, `anex`, and `emex`, and answer any of the TA's questions. The lab must be demonstrated before the end of the lab period on September 25, 2009; otherwise, the demonstration will be marked as "incomplete".

Evaluation

The following three components of this lab will be evaluated separately:

- Written work and plots;
- MATLAB code; and
- Lab demonstration.

Each component is weighted equally, and graded on the following five-point scale:

- 5: Outstanding work demonstrating original thinking
- 4: Satisfies the lab requirements
- 3: Minor issues in satisfying the lab requirements
- 2: Major problems in satisfying the lab requirements
- 1: Work is incomplete
- 0: Work is missing (or student is absent for the demonstration)

Note that the maximum grade for satisfying the basic lab requirements is an "A" (80%).