# Homework #03 • MATH495/STAT490 • Great Expectations

- submit your write-up Wednesday 01 October.
- participation in webct discussions is encouraged.
- page limits will be enforced.
- highlight major results.

**A) Expected Value of a Function** (3 pages, 10 pts) You are aware that the expected value of a function (of a random variable) is NOT the function evaluated at the expected value (mean). In some cases, however, one can approximately quantify the difference. For a random variable $x$ with $\text{mean}[x] = \mu$ and $\text{var}[x] = \sigma^2$, consider the random-valued function $h(x)$. When the variance is small ($\sigma^2 \ll 1$), it turns out to be a reasonable approximation to replace $h(x)$ by its quadratic Taylor expansion about $x = \mu$ in the expected value integral

$$E[h(x)] = \int_{-\infty}^{+\infty} h(\tilde{x})\, f(\tilde{x})\, d\tilde{x} \ .$$

- Explain in 1-2 sentences why you might think this to be a reasonable approximation.
- Express the approximate expected value, $E^a[h(x)]$, in terms of $\mu$ and $\sigma^2$.
- Use a similar approximation to show that $\text{var}[h(x)] \approx (h'(\mu))^2\, \sigma^2$.
- The Matlab simulation *w04fun.m* produces estimates of the mean and variance for $h(x) = e^{-x}$ where $x$ is normally distributed with $\mu = 5$ and given $\sigma^2$. Does the sign of the difference $E^a[h(x)] - h(\mu)$ make sense? By carefully observing a series of simulations, see how well these approximations work – you may find the results a bit surprising. Can you give an explanation?

**B) Means & Variances** (3 pages, 10 pts) Problems #37, #52a and #56 from Chapter 2 of Ross. Calculate the variance for the random process of #37. Note that #56 is closely based on one of the examples in the chapter.

**C) Random Simulation** (2 pages, 10 pts) Modify the lecture demo *w04exp.m* to simulate a continuous random variable whose probability density function (PDF) is

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 2x\, e^{-x^2} & \text{for } 0 \leq x \end{cases}$$

Produce a sample empirical (cumulative) distribution function.

However, if you are given the above empircal CDF, producing the corresponding probability density function is tricky business, as it involves numerical differentiation. To see why this is so, try the following. Use as the data points for the differentiation $(x_j, j/n)$ for $j = 0 \to n$ where the $x_j$ are the sorted random data. For the $j = 0$ index, take $x_0 = 0$. For uniformly distributed points, a derivative can be approximated by the (centered) finite difference

$$f'(x) \approx \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x} \ .$$

However, our $x_j$'s are not uniformly distributed — but our vertical coordinates are! So, the reciprocal slopes can be obtained by the above finite difference formula. Discuss your (likely disappointing) results. Any ideas for improving them?