# Introduction to Stochastic Differential Equations
# MATP 4960/6960–1 – Spring 2007
# Homework 2

Due Thursday, October 18 at 5 PM

This homework has 160 points for MATP 6960 (210 points for MATP 4960) available but, as always, homeworks are graded out of 100 points. Full credit will generally be awarded for a solution only if it is both correctly and efficiently presented using the techniques covered in the lecture and readings, and if the reasoning is properly explained. If you used software or simulations in solving a problem, be sure to include your code, simulation results, and/or worksheets documenting your work. If you score more than 100 points, the extra points do count toward your homework total.

## 1 MATP 4960 Problems

### 1.1 Value Coverage (10 points)

Suppose that $X_1, \ldots, X_n$ are $n$ independent, identically distributed random variables drawn from a common probability distribution with CDF $F_X$. Derive an expression for the probability that a specified value $x$ lies strictly between the minimum and maximum values of the $\{X_1, \ldots, X_n\}$.

### 1.2 Failure of Redundant System (10 points)

Consider an electronic system consisting of $n$ redundant components, each of which has a lifetime (until failure) which can be modeled as a random variable independent of the lifetimes of the other components. The system operators wait until $k$ components have failed before they take the system offline and replace all components with new ones. Therefore, the components in operation always have started at the same time. Derive an expression for (some quantitative representation of) the probability distribution for the amount of time the system operates before the system is taken offline for component replacement, in terms of some quantitative representation of the probability distribution of the lifetime of each component.

### 1.3 Gaussian Race (30 points)

Implement a random number generator for standard Gaussian (normal) random variables (mean zero and variance one) based on the Box-Muller method.

a. (**15 points**) Demonstrate the validity of your algorithm by comparing histograms of simulated Gaussian random variables against the theoretical probability distribution.

b. (**10 points**) Develop another Gaussian random variable simulator based on a direct implementation of the Inverse Transform method, and show that it is working properly by comparing simulated histograms against the theoretical probability distribution.

c. (**5 points**) Compare how fast your code can simulate a given number $n$ of Gaussian random variables when you use the Box-Muller method as compared to the straightforward Inverse Transform method. Try a few large values of $n$, and report the times required for the simulations and the times required per random number simulated. Comment on the behavior of the algorithms for not-so-large and for large values of $n$.

## 2 MATP 4960/6960 Problems

### 2.1 Mean Quotients (10 points)

Given independent random variables $X$ and $Y$ with absolutely continuous probability distributions concentrated on the positive real numbers, derive a formula for the mean value of the quotient $Q = X/Y$ in terms of the probability densities of $X$ and $Y$. Under what conditions is this expression finite or infinite?

### 2.2 Deviation from Median (20 points)

Suppose an odd number, $n$, of independent identically distributed random variables is sampled from a common absolutely continuous probability distribution. Derive a quantitative expression for the difference between the maximum value sampled and the median (middle value) of the samples in terms of some quantitative descriptors of the probability distribution which generated the random variables. This quantity could be of relevance in determining (for example in socioeconomic studies) how much the extreme outliers fall from the middle of the pack purely due to chance under fair circumstances.

## 2.3 Exponential Independence (30 points)

Suppose that a server is awaiting two tasks which are being preprocessed (in parallel) by another system which received the tasks simultaneously. The preprocessing times for the two tasks can be modeled as a pair of independent random variables with a common probability distribution. The server that will receive the tasks after the preprocessor must handle them one at a time, so the manager is interested in knowing the statistical properties of the following two variables:

- The time the server must wait until the preprocessor finishes and passes on either one of the tasks to the server,
- The time the server has available to work on the first task it receives from the preprocessor until it receives the second task.

a. (**10 points**) Compute the joint and marginal probability distributions (in whatever quantitative representation you like) of these two time variables if the preprocessing time is exponentially distributed.

b. (**20 points plus bonus**) Show that the exponential distribution is the only absolutely continuous probability distribution for the preprocessing time which makes the two times described above behave as independent random variables. For bonus credit, extend your argument to show that the exponential distribution is the only probability distribution (including those which are not absolutely continuous) which has this property.

## 2.4 Feel the Force (20 points)

Consider a simple model of a stellar cluster: a collection of $n$ stars distributed within a spherical region of radius $R$, each star with the same mass $M$. The gravitational vector field exerted by an object of mass $M$ located at position $\mathbf{x}_{\mathrm{s}}$ on another object at position $\mathbf{x}$ is

$$\mathbf{F}(\mathbf{x}) = \frac{GM(\mathbf{x}_{\mathrm{s}} - \mathbf{x})}{|\mathbf{x}_{\mathrm{s}} - \mathbf{x}|^3},$$

where $G$ is a universal gravitational constant. Suppose the $n$ stars are each independently and uniformly distributed within the region bounded by a sphere of radius $R$.

Fix an arbitrary unit vector $\hat{\mathbf{e}}$ and compute the probability distribution for the gravitational vector field component along $\hat{\mathbf{e}}$ at the center of the sphere exerted by *one* of the $n$ stars.

We will later return to this problem and use the result you obtain as a basis for computing the gravitational force induced by adding up the forces from all $n$ stars in the cluster.

## 2.5 Trains in Vain (20 points)

Imagine that the capital district had a light rail system, with one "Past Glory Transit" going from Troy through Latham to Schenectady, and another "Mall Metro" starting in Saratoga Springs, running south through Latham to Guilderland. Suppose the transit designers realized that quite a few people would like to go from Troy to Guilderland (Crossgates Mall, etc.) and so they arrange the train schedules so that the Mall Metro is supposed to arrive in Latham from Saratoga at the same time the Past Glory Transit arrives in Latham from Troy. Each train would then wait 2 minutes in the station to unload and load passengers, hopefully leaving enough time for everyone except Andrew to change trains. Unfortunately, the operators of this light rail system take a cue from Amtrak (or British Rail) and consider the train schedule as merely a "suggestion" and the trains may arrive rather late. The light rail operators are considering trying to coordinate the trains so that the Mall Metro will always wait in Latham so that the people arriving in Latham on the Past Glory Transit from Troy have exactly 2 minutes to board the train before it leaves for Guilderland. In particular, that means the Mall Metro will wait in the Latham station for more than 2 minutes if it arrives from Saratoga before the Past Glory Transit arrives from Troy. The managers of the light rail system would like to know what sort of statistics to expect for the amount of time the Mall Metro would wait in Latham under such an arrangement.

Suppose statistics are available for the fraction of time each train arrives on time as well as for the length of delays when they are late. You can neglect the case of trains arriving early (these would be treated as on-time and the managers don't care about the amount of extra time they are waiting in the station from being early).

a. (**7 points**) Construct a probability model, indicating how the statistics about on-time arrival percentages and delay times would be incorporated in your mathematical framework. You can either illustrate by example or a simple general explanation (which can be brief and precise). You may assume that delays on each train line are independent of each other.

b. (**13 points**) Use your mathematical framework to compute the probability distribution for the amount of time the Mall Metro waits in the Latham station. You can use any quantitative representation to describe the probability distribution which you find convenient.

## 2.6 The Answer, My Friend, is Cyclin' in the Wind (30 points)

In a high-stakes outdoor bicycle race, a cycling team's trainer is studying the cycling course and trying to determine what kind of wheels the bicycles should have and how the cyclists should pace themselves. One element of his analysis is the drag force the cyclists will experience (in the absence of drafting behind other racers, a lá Andy Bernard and Kevin in the "Fun Run" episode). The magnitude of the drag force experienced (assuming the cyclists are moving fast enough to be at high Reynolds number) can be expressed as

$$F = \gamma |\mathbf{V} - \mathbf{W}|^2$$

where $\gamma$ is some drag coefficient, $\mathbf{V}$ is the vector-valued velocity of the bicycle, and $\mathbf{W}$ is the vector-valued velocity of the wind. (Actually the drag coefficient $\gamma$ should depend on the relative velocity of the bicycle to the wind, since the bicycle and cyclist present different profiles along different directions, but we neglect this here for simplicity.) All vectors here are along the two-dimensional horizontal plane; we neglect vertical components of the bicycle and wind velocities.

Suppose the bicyclist velocity $\mathbf{V}$ is to be specified by the trainer (even if it means the cyclist has to pedal harder against a headwind), but the windspeed $\mathbf{W}$ is unknown ahead of the race. The trainer would like to know the probability distribution for the drag the cyclist would feel if he were ordered to ride this segment of the course at velocity $\mathbf{V}$.

a. (**15 points**) Suppose the direction of the wind is uniformly distributed, the magnitude of the wind velocity is given by a known absolutely continuous probability distribution supported on the nonnegative real numbers, and the wind magnitude and direction can be assumed independent of each other. Construct a probability model under these assumptions to describe the probability distribution (using whatever quantitative representation you like) for the drag force felt by the bicyclist.

b. (**5 points**) Provide a hypothetical but realistic situation in which the assumptions in the previous part might be violated.

c. (**10 points**) Suppose we now use a more general and detailed probability model based on a given probability density function $p_{\mathbf{W}}$ for the wind velocity vector $\mathbf{W}$, without the simplifying assumptions of part a. We still are assuming the probability distribution for the wind velocity to be absolutely continuous on $\mathbb{R}^2$. Derive a quantitative expression for some quantitative representation of the probability distribution for the drag force in terms of the given $p_{\mathbf{W}}$.

## 2.7 A Perfect Foam (30 points)

Suppose that a plane is scheduled to fly in a straight line at constant speed $v$ for a distance $\ell_F$ through a stormy region. The airline wants to cancel the flight if the chance that the plane will encounter three or more storms (on its scheduled flight path) is greater than 50%. To determine whether to cancel the flight, the airline analyst wishes to use a Poisson point process model in space and time for the storms.

To prepare, one must first recognize that it doesn't make sense to define a Poisson point process directly in space and time since they have different dimensions, and there is no reason the intensity along the space directions should be the same as the intensity along time directions. (More precisely the intensity along space and time directions depends on the choice of reference units used to measure space and time). Consequently, we nondimensionalize space with respect to a length scale $\ell_C$ and time with respect to a time scale $\tau_C$ that are characteristic of intervals of space and time, respectively, between storms. Taking $(x, y)$ as the original horizontal spatial coordinates and $t$ as the original time variable, we define nondimensionalized variables $x' = x/\ell_C$, $y' = y/\ell_C$, and $t' = t/\tau_C$, and now define our storm model as follows:

- Storm centers are initiated at points distributed according to a Poisson point process model with intensity 1 in the three-dimensional space of nondimensional space-time variables $(x', y', t')$.
- Each storm remains fixed in space, affects a circular region of radius $\ell_S$ (therefore $\ell_S/\ell_C$ in the nondimensionalized space variables $(x', y')$) centered at the Poisson point, and remains active for a time $\tau_S$ (and so for an interval of length $\tau_S/\tau_C$ in terms of the nondimensionalized time variable $t'$) after the Poisson point initiating it.

We completely neglect the vertical position of the plane and storms here. All questions will be phrased in terms of the original dimensional space and time variables, so to compute with the Poisson point process model, be careful to make the change of variables back and forth from the nondimensionalized variables in terms of which the Poisson point process model is posed.

a. (**10 points**) Show that this nondimensionalized model makes sense by verifying the following two statements:

   (a) (**5 points**) The number of storms initiated with centers in a horizontal region $B$ of area $\ell_C^2$ over an arbitrary time interval of length $t$ has mean $t/\tau_C$.

   (b) (**5 points**) The number of storms initiated over a horizontal region $B$ of arbitrary area $A$ during a given time interval of length $\tau_C$ has mean $A/\ell_C^2$.

b. (**20 points**) Now use the probability model to determine a quantitative criterion for canceling the flight in terms of the storm parameters ($\ell_C$, $\tau_C$, $\ell_S$, and $\tau_S$) and flight path properties ($\ell_F$ and $v$). Note that all these parameters are given in ordinary dimensional terms.