Advanced Probability Modeling and Techniques
MATH 4960/6960 – Fall 2007
Homework 1

Due Friday, September 21 at 2 PM

This homework has 155 points for MATP 6960 (200 points for MATP 4960) plus 45 bonus points 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 Out of Committee (15 points)

Suppose in a group of 8 people, a committee of three members is selected in the first year, a committee of four members is selected in the second year, and a committee of five members is selected in the third year. Each of these committees is supposed to be chosen at random with no favoritism, and membership on a committee one year does not alter the probability for being chosen to serve on a committee in a future year. One of the group of 8 people was chosen for none of the committees and claims that there must be some deliberate manipulation of the random selection process. In an ensuing civil court case, a lawyer hires you as a mathematical consultant to provide some hard numbers.

a. (**5 points**) If the committee selection process were truly fair, what would be the probability that the particular person in question would be chosen for none of the committees in the three years?

b. (**10 points**) If the committee selection process were truly fair, what would be the probability that each person in the group would be selected for a committee at least once in the three years?

1.2 Raisin Cookies (10 points)

Suppose you are running a snack factory and are designing a new line of raisin cookies. Assuming the raisins just get mixed in a big batch of dough before being separated into cookie pieces, estimate the average number of raisins per cookie you would need so that at least 99% of the cookies have raisins. Explain your reasoning.

1.3 Line Noise (10 points)

Suppose bits of information, with values 0 or 1 are sent over a transmission line, but are subject to distortion, which we model by adding a normally distributed random variable with mean 0 and standard deviation σ to the value of 0 or 1 that was intended to be transmitted. (Of course some measurable quantity like signal amplitude or frequency is altered rather than a literal value of 0 or 1 in the line, but the idea is the same.) Suppose that the receiver assumes the value 0 was transmitted if he receives a value less than $1/2$, and assumes the value 1 was transmitted if he receives a value greater than or equal to $1/2$. What is the probability that the receiver makes an error in interpreting the transmitted bit? Express your answer in terms of σ and as always, explain your reasoning, but you do not need to simpilfy your answer.

1.4 Absolut Probability (10 points)

Suppose a random variable X is uniformly distributed on the real interval $[a, b]$. Show that $|X|$ has an absolutely continuous probability distribution and derive its probability density.

2 MATP 4960/6960 Problems

2.1 Another Random Variable Living on the Square (15 points)

Given the probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = [-1,1]^2$, $\mathcal{F} = \mathcal{B}(\Omega)$, and P is the uniform probability measure on Ω, define the random variable $X : \Omega \to \mathbb{R}$ so that $X(\omega)$ is the distance of ω from the origin.

a. (**7 points**) Describe clearly the σ-algebra $\sigma(X)$ generated by the random variable X, preferably through pictures and precise English sentences.

b. (**3 points**) Provide an example of a set A satisfying $A \in \mathcal{F}$ but $A \notin \sigma(X)$.

c. (**5 points**) Provide explicit formulas for the probability distribution induced by X on open intervals (a, b) in $\mathbb{R}$.

2.2 Slime Creatures from Product Space (20 points)

Consider a probability space $(\Omega, \mathcal{F}, P)$ where the sample space $\Omega = [0,1]^\infty$ consists of infinite sequences of real numbers in the unit interval, $\mathcal{F} = (\mathcal{B}([0,1]))^\infty \equiv \mathcal{B}([0,1]) \times \mathcal{B}([0,1]) \times \cdots$ is the product σ-algebra generated by finite-dimensional rectangles (or cylinders), and $P = P_1 \times P_2 \times \cdots$ is a product measure with all $\{P_j\}_{j=1}^\infty$ being uniform probability measures on $[0,1]$.

a. (**15 points**) Represent a point $\omega \in \Omega$ as a sequence $\omega = (\omega_1, \omega_2, \ldots,)$ with $\omega_j \in [0,1], j = 1, 2, \ldots$. Define $X(\omega) = \sup_{j \geq 1} \omega_j$. Explain whether or not $X(\omega)$ is a random variable on the probability space $(\Omega, \mathcal{F}, P)$. If it is a random variable, explain the structure of the σ-algebra $\sigma(X)$ generated by X.

b. (**5 points**) How do your answers change if the P_j are probability measures on $[0,1]$ which are not necessarily the uniform probability measure?

2.3 Sets in Independent σ-Fields (5 points)

Given a probability space $(\Omega, \mathcal{F}, P)$ with independent σ-algebras $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$, show that $A \in \mathcal{F}_1 \cap \mathcal{F}_2$ implies that either $P(A) = 0$ or $P(A) = 1$.

2.4 Failure of Probability Extension (20 bonus points)

Consider the collection of sets $\mathcal{C} = \{(0, 1/n), [1/n, 1]\}_{n=1,2,3,\ldots}$ in the sample space $\Omega = (0,1]$. Define a set function P on $\mathcal{C}$ so that $P((0,1/n)) = 1$ and $P([1/n,1]) = 0$

for all $n = 1, 2, \ldots$.

a. (**15 bonus points**) Prove that P cannot be extended to a probability measure on $\sigma(\mathcal{C})$, the σ-field generated by $\mathcal{C}$.

b. (**5 bonus points**) Explain carefully and precisely why this result is not in contradiction with the probability extension theorem stated in, for example, Grigoriu, *Stochastic Calculus*, Sec. 2.3.3.

2.5 Probability Sphere (20 points)

Consider a probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = S^1 = \{\mathbf{x} \in \mathbb{R}^3 | |\mathbf{x}| = 1\}$ is the unit sphere in three dimensions, $\mathcal{F} = \mathcal{B}(S^1)$, and P is the uniform probability measure. Define random variables Θ and Φ on this probability space as the values of the usual (poloidal and azimuthal) spherical angles, that is, so that the random variables corresponding to the Cartesian coordinates X_1, X_2, and X_3 are expressed:

$$\begin{aligned} X_1 &= \sin\Theta\cos\Phi, \\ X_2 &= \sin\Theta\sin\Phi, \\ X_3 &= \cos\Theta. \end{aligned}$$

a. (**5 points**) Derive the CDF for Φ, and if the probability distribution is absolutely continuous, provide the probability density.

b. (**8 points**) Derive the CDF for Θ, and if the probability distribution is absolutely continuous, provide the probability density.

c. (**7 points**) Determine whether or not Φ and Θ are independent random variables, and justify your conclusion.

2.6 Chromosomal Square Dance (20 points)

Consider a cell with N chromosomes (need not be human!) which is subjected to radiation which causes some of the chromosomes to break and recombine with each other. In particular, we define a chromosomal interchange to be an event in which two chromosomes break, exchange one of their pieces with each other, and each recombine (the wrongly matched pieces). Suppose r chromosomal interchanges occur, and that each chromosome is equally likely to be involved in a chromosomal interchange. What is the probability that exactly m of the N chromosomes will be corrupted due to an interchange? You may ignore the possibility of a chromosomal interchange being "undone" by just the right rebreaking and recombination.

2.7 Intel Idol (20 points)

Suppose a parallel computing algorithm delegates three tasks to three separate processors, to which we will refer as A, B, and C. Each processor starts its task at the same time and performs its computation in parallel, except near the end of its computation, processor B requires the result from processor A, and similarly, near the end of its computation, processor C requires the result from processor B. Therefore, the parallel algorithm will proceed with minimal waste of processor time if the processors finish their main computations in the order: A then B then C. By "main computation" we mean the computations up to the point that a processor needs a result from another processor. The times required by each processor for their main computation are rather uncertain, however, due to variability in the computational task to be performed and the load on the processors due to other computations. Suppose that, from some statistical analysis, we model processor A as requiring a uniformly distributed amount of time varying between 1 and 2 seconds for its main computation, and processor B as requiring an exponentially distributed amount of time with mean 2 seconds for its main computation. Processor C's main computation is actually two subtasks which are performed in succession (one right after the other), each of which requires a time which is exponentially distributed with mean 1.5 seconds. Under such a probability model, what fraction of the time will the processors complete their main computations in an order other than the desired one (A then B then C)?

2.8 Ringing Risk (15 points)

In electronic and mechanical control systems such as shock absorbers in vehicles, it is sometimes desirable to introduce a critically damped oscillator. A linear mechanical oscillator can be characterized by mass m, force constant k, and damping coefficient γ; and its response to disturbances characterized by its eigenvalues which are the roots of the equation

$$m\lambda^2 + \gamma\lambda + k = 0.$$

(Similar mathematical equations apply for electronic circuits, though different symbols are typically used.) The system is critically damped when the eigenvalues have negative real part and no imaginary part, which happens for example when the system is designed with $m = k = 1$ and $\gamma = 2$ (in suitably nondimensionalized units). However, in the manufacturing of the circuit, these parameters have uncertainties. For simplicity, suppose m is uniformly distributed between $1 - \delta_m$ and $1 + \delta_m$, k is uniformly distributed between $1 - \delta_k$ and $1 + \delta_k$, and γ is uniformly distributed between $2 - \delta_\gamma$ and $2 + \delta_\gamma$, where δ_k, δ_m, and δ_γ are certain nondimensionalized uncertainty parameters. Suppose that undesirable ringing occurs when an eigenvalue has an imaginary part which is greater in magnitude than some critical value Ω. Under the above model, express the probability for undesirable ringing in the circuit in terms of the uncertainty parameters and Ω.

2.9 Assessment of Quality of Generation of Random Numbers (40 points)

Use the methods described in class to write programs which generate independent random numbers with the following probability densities:

a.
$$p_X(x) = \begin{cases} \frac{4}{3}x^4 e^{-2x}, & x \geq 0, \\ 0 & x < 0. \end{cases}$$

b.
$$p_X(x) = \begin{cases} \frac{1}{\pi\sqrt{x(1-x)}}, & 0 < x < 1, \\ 0 & x \leq 0 \text{ or } x \geq 1. \end{cases}$$

For full credit, your algorithm should make use of any efficiency-enhancing or simplifying techniques discussed in class or the assigned reading. Be sure to include a copy of your code used to generate your results.

2.9.1 Generation of Histograms (15 points)

For each case, generate a large number (like 10^4) independent samples of the random variable and plot them on a histogram. For random variables with an infinite range, a sensible thing to do is to choose some moderate sized finite interval and divide it into equally spaced intervals. Then one defines bins for each of these intervals, as well as one for values which fall to the left of the finite interval and one for values which fall to the right of the finite interval. The finite interval should be wide enough so the number of samples which fall in the outside bins is relatively small compared to those falling in the internal bins.

2.9.2 Comparison Against Correct Probability Distribution (15 points)

Compute (numerically if necessary) for each of the random variables the probability that the random variable should fall into each of the bins, if it's being properly simulated. Call this the "theoretical" binned probability distribution. Compute your "simulated" binned probability distribution by dividing the number of times your random variable fell into a given bin by the number of random variables you simulated. Plot the "theoretical" and "simulated" binned probability distributions on the same graph. How do they compare?

2.9.3 Comparison Against Correct Mean and Variance (10 points)

To make the comparison more quantitative, compute the *sample mean*

$$\hat{\mu} = \frac{1}{N} \sum_{j=1}^{N} X_j$$

and *sample variance*

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{j=1}^{N} (X_j - \hat{\mu})^2$$

where N is the number of random variables you simulate and X_j are the realizations of the random variable. Compare these sample means and variances to the correct theoretical values. Be sure to explain how you obtain the theoretical values.

You may be puzzled by why the sample variance has $N - 1$ rather than N in the denominator. The reason is statisical: dividing by N would produce an underestimation bias of the variance. That's because you're computing the fluctuations from your *sample* mean, but your *sample* mean is itself a random variable with variance σ^2/N. To correct for this (since you're trying to estimate the mean-square fluctuations with respect to the *true* mean μ), the denominator for $\hat{\sigma}^2$ is made to be $N - 1$.

2.10 Discrete and Continuous Probability can be Handled within Measure-Theoretic Framework (25 bonus points)

Sometimes it is helpful to consider a single continuous state space (such as $S = \mathbb{R}^d$) where both discretely and continously distributed random variables can live. One does this by associating to each random variable $\mathbf{X}$ a measure $P_{\mathbf{X}}$ so that

$$\mathrm{Prob}\,(\mathbf{X} \in B) = P_{\mathbf{X}}(B)$$

for all Borel sets $B \subseteq S$. Then averages involving the random variable $\mathbf{X}$ are expressed as

$$\langle \mathbf{f}(\mathbf{X}) \rangle = \int_S \mathbf{f}(\mathbf{x})\, \mathrm{d}P_{\mathbf{X}}(\mathbf{x}) \qquad (1)$$

with the integral interpreted in the sense of Lebesgue (as a limit where $\mathbf{f}$ is approximated by "simple functions"). See Billingsley's *Probability and Measure* or Folland's *Real Analysis* for a decent discussion and other texts for not so nice discussions.

We discussed in class the case of an (absolutely) continuously distributed random variable, for which there exists an (integrable) probability density $p_{\mathbf{X}}$ so that

$$P_{\mathbf{X}}(B) = \int_B p_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{2}$$

Then

$$\langle \mathbf{f}(\mathbf{X}) \rangle = \int_S f(\mathbf{x})\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}) = \int_S \mathbf{f}(\mathbf{x}) p(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

can be interpreted in the less intimidating Riemann-Stieltjes sense (limit of a fine discretization of the independent variable $\mathbf{x}$) if $\mathbf{f}$ and $p_{\mathbf{X}}$ are piecewise continuous.

In this problem, consider a random variable $\mathbf{X}$ which only takes a finite or countably infinite set of values $\{\mathbf{a}_n\}_{n\in I}$ with

$$\mathrm{Prob}\,\{\mathbf{X} = \mathbf{a}_n\} = p_n$$

and $\sum_{n\in I} p_n = 1$. The twist is that rather than taking the state space to be $\{\mathbf{a}_n\}_{n\in I}$, we take it to be $S = \mathbb{R}^d$. You are asked to show this can be done within the general measure-theoretic framework.

To do this, define the measure

$$P_{\mathbf{X}} = \sum_{n\in I} p_n \delta_{\mathbf{a}_n}$$

where the *Dirac measure* $\delta_{\mathbf{x}}$ is defined:

$$\delta_{\mathbf{x}}(B) = \begin{cases} 1 & \text{if } \mathbf{x} \in B, \\ 0 & \text{if } \mathbf{x} \notin B. \end{cases}$$

for Borel sets $B \subseteq S$. Note that the measure $P_{\mathbf{X}}$ is certainly *not* absolutely continuous (with respect to Lebesgue measure), so that one cannot write an expression like (2) for an ordinary function $p_X(\mathbf{x})$.

Prove that the general measure-theoretic definition of the statistical average (1), interpreted as a Lebesgue integral, recovers the familiar result from discrete probability:

$$\langle \mathbf{f}(\mathbf{X}) \rangle = \int_S \mathbf{f}(\mathbf{x})\,\mathrm{d}P_{\mathbf{X}}(\mathbf{x}) = \sum_{n\in I} p_n \mathbf{f}(\mathbf{a}_n),$$

provided $\mathbf{f}$ is a continuous function.

What can go wrong if $\mathbf{f}$ is discontinuous?