Due Date: This assignment is due on Friday, 8 May, 2009(with the natural 1 day extension for SCPD students), by 5pm in the box outside Durand 112. The LaTeX incentive policy is in effect.

Problem 1: Consider Monte Carlo integration with $g \geq 0$ and assume that $g \leq ch$ for some $c > 1$ and some density h. The *hit-or-miss* estimator of $z \stackrel{\text{def}}{=} \int_0^1 g(u)du$ is $c\mathbb{1}_{\{Uch(Y) \leq g(Y)\}}$, where U, Y are independent with U as uniform$(0,1)$ and Y with density h (note that h could be defined over the whole real line, but we restrict this to h only defined on $[0,1]$). Show that its expectation is $z = \int_0^1 g(u)du$ as desired, but that the variance is always at least the variance of the importance sampling estimator that uses sampling from h.

Problem 2: With $\phi(y)$ as in Problem 7 from Homework 1, show the following.

1. Prove that, for all functions g satisfying $\mathbf{E}[g(Y)^2] < \infty$, $\mathbf{E}[(X - \phi(Y))g(Y)] = 0$.
2. Prove that $\phi(Y)$ is the best mean square predictor of X across all predictors $g(Y)$ such that $\mathbf{E}[g(Y)^2] < \infty$.

Problem 3: Variance Reduction with Control Variates
Suppose that we wish to compute $\alpha = \mathbf{E}X$ via Monte Carlo. Assume that there exist r.v.'s $Y_1, \ldots, Y_d$ such that $\mathbf{E}Y_i = \mu_i$ is known for $1 \leq i \leq d$. The r.v. $C_i = Y_i - \mu_i$ is called a *control variate*.

1. If $C = (C_1, \ldots, C_d)^T$, prove that
$$\mathbf{E}X(\lambda) = \alpha,$$
where $X(\lambda) = X - \lambda^T C$, $\lambda \in \mathbb{R}^d$.
2. Find the vector λ^* which minimizes $\operatorname{Var}(X(\lambda))$ over λ.
3. How would you estimate λ^* in an implementation of this approach?
4. How general is this method? (i.e., are there typically r.v.'s for which the means $\mu_1, \ldots, \mu_d$ can be computed?) Explain your answer.

Problem 4: The expected shortfall of a rv Z at the p'th quantile is defined as

$$\mathbf{E}\left[Z | Z > q\right]$$

where q is the p'th quantile value, that is $\mathbf{P}\left\{Z \leq q\right\} = p$.

This problem is concerned with estimating the expected shortfall of a so-called *Asian Option*. The option is based on an underlying asset V which evolves in discrete time according to

$$V_{n+1} = V_n R_{n+1}$$

where R_{n+1} is a log-Normal random variable, i.e.

$$\ln R_n \sim \mathrm{N}\left(r, \sigma^2 \Delta t\right),$$

r is the risk free rate of return over one period, and σ^2 is the volatility of V. The price of a (simplified) Asian option with *expiration* N time units in the future is

$$X_N = \mathbf{E}_{V_0}\left(N^{-1}\sum_{i=1}^{N} V_i - K\right)_+$$

where K is the *strike price*. In general, an analytic expression is not known for X_N. (Note: $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ if $x < 0$.)

We will treat the V_n's as the weekly values of the asset. The annual risk-free rate of return is 5%, the volatility $\sigma^2 = 0.1$, $N = 50$ weeks[1] , $V_0 = \$100$ and $K = \$115$. (Note: Since the risk-free rate and σ^2 were given as annual values, the generation of R_n must reflect that. The volatility in R_n was given as $\sigma^2 \Delta t$, where $\Delta t = \frac{1}{50}$, so it already adequately reflects this. For the weekly return, the effective rate is $r = \ln(1.05)/50 = 9.7580 \times 10^{-4}$.)

Using a bootstrap technique, compute the 95% confidence interval for the expected shortfall of the rv

$$\left(N^{-1} \sum_{i=1}^{N} V_i - K \right)_+ ,$$

with $p = 90\%$. Please include a description of the your method as well as the code.

Problem 5: Suppose that $x(\cdot)$ is the solution to a deterministic differential equation

$$\frac{d}{dt} x(t) = \phi(\theta, x(t))$$

such that

$$x(0) = x_0$$

where ϕ is deterministic and θ represents a vector of parameters. (For example, $\phi(\theta, x) = \theta x$ in Example 13 on page 52.) Assume that x_0 and θ are measured with error, and that $(X_0, \hat{\theta})^T$ are multivariate normally distributed with mean $(x_0, \theta)^T$.

1. Compute the small noise approximations for the solution $X(t)$ to

$$\frac{d}{dt} X(t) = \phi(\hat{\theta}, X(t))$$

such that

$$X(0) = X_0$$

2. Discuss the computational issues that arise in computing the variance of your small noise approximation

Problem 6: Develop a corresponding approximation confidence interval for $q(p)$, where $q(p)$ is the "p^{th} quantile" of the random variable Y defined as the smaller root of the equation

$$P\{Y \leq q(p)\} = p$$

Such quantile computations are of interest on "value at risk" calculations in the finance setting.

Problem 7: Let X be a Cauchy r.v. so that its density is given by

$$f(x) = \frac{1}{\pi(1 + (x - b)^2)}$$

1. Compute the distribution of $X_1 + X_2$ where X_1, and X_2 are independent copies of X.
2. If $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from X, what does $n^{-1}(X_1 + \cdots X_n)$ converge to?
3. How might you estimate the parameter b (in lieu of part 2)?
4. Explain how to generate the r.v. X using inversion.
5. Consider the following algorithm:

[1] There are 50 working weeks in the financial year.

i. Generate independent r.v.'s V_1 and V_2 that are uniform on $[-1, 1]$.

ii. If $V_1^2 + V_2^2 \leq 1$, return $X = V_2/V_1$; else, return to step i.

Show that X is Cauchy distributed. (This can be faster than inversion when computing the arc tangent is expensive.)

Problem 8: Consider a linear model in which

$$Y_i = a^* x_i + b^* + \epsilon_i, \quad 1 \leq i \leq n,$$

where the ϵ_i's are iid rvs with density

$$f(x) = \frac{\lambda^*}{2} e^{-\lambda^* |x|}.$$

Note: From a terminology standpoint, it is important to recognize that this is different from a linear regression problem. A linear regression problem aims to find a least squares solution for the model. In this problem, we are looking for maximum likelihood estimates for the model. These are distinct problems, except when the error is normally distributed. In that case, they are equivalent.

1. The MLE for a^*, b^* and λ^* solves an optimization problem. What is it?
2. Show that the MLE can be computed as the solution to a linear program.

Problem 9: Least Squares Estimation

Name	Height (in)	Weight (lbs)	Age
Alfred	69.0	112.5	14
Alice	56.5	84.0	13
Barbara	65.3	98.0	13
Carol	62.8	102.5	14
Henry	63.5	102.5	14
James	57.3	83.0	12
Jane	59.8	84.5	12
Janet	62.5	112.5	15
Jeffrey	62.5	84.0	13
John	59.0	99.5	12
Joyce	51.3	50.5	11
Judy	64.3	90.0	14
Louise	56.3	77.0	12
Mary	66.5	112.0	15
Philip	72.0	150.0	16
Robert	64.8	128.0	12
Ronald	67.0	133.0	15
Thomas	57.5	85.0	11
William	66.5	112.0	15

Table 1: Data For Problem 9

1. Use the first two columns of data (Height and Weight) in Table 1 to build various regression models that attempt to explain weight as a function of height. Which of your models does the best job of predicting the weights of individuals in the third column? (There is no need to use the age here. If you wanted to, there is a way to factor in the age.)
2. Another possibility is to build separate regression models by gender. Does this improve the predictions?

3. Now use the full data set to recompute the coefficients for the type of regression model that worked best in part a. Construct a confidence interval for the slope based on bootstrapping. (Take into account the fact that both weight and height should be viewed as random variables in this setting. In other words, this is not a setting where the levels of the explanatory variable (in this case, height) is carefully set by the experimenter at various predetermined levels; the height values that are observed are determined by the particular random sample that is selected.)

4. Suppose that the height of a student is 71 inches. Use the bootstrap to construct a 95% prediction interval for that student's weight.

Problem 10: When the arrival process to a single-server queue follows a so-called Poisson process having rate λ (i.e. the inter-arrival times $\chi_1, \chi_2, \dots$ are i.i.d. exponential random variables having parameter $\lambda > 0$) with $\rho = \lambda \mathbf{E} V_0 < 1$, the steady-state random variable W_∞ corresponding to the time spent waiting in the queue can be represented as

$$W_\infty = \sum_{i=1}^{N} Z_i$$

where N is a geometric random variable having probability mass function $\mathbf{P}(N = k) = (1-\rho)\rho^{k-1}$ and $(Z_i : i \geq 1)$ is an i.i.d. sequence of random variables independent of N satisfying

$$\mathbf{P}(Z_1 > x) = \frac{1}{\mathbf{E} V_0} \int_x^\infty \mathbf{P}(V_0 > y) dy$$

1. Suppose that V_0 is exponential with parameter $\mu > 0$. Compute the distribution of W_∞.

2. Abandoning the assumption in part (a), write W_∞ as $W_\infty(\lambda)$ (reflecting its dependence on λ). Prove that if $\mathbf{E} V_0^2 < \infty$, $(1-\rho) W_\infty(\lambda) \Rightarrow \Gamma$ as $\lambda \to \frac{1}{\mathbf{E} V_0}$ and compute the distribution of Γ (Γ is not necessarily a Γ distributed random variable. We just needed a letter and it it seemed as good as any other).

   Hint: Convergence in distribution is equivalent to point-wise convergence of characteristic functions. Use that fact to show the desired convergence as well as to derive the limiting characteristic function, and thus the distribution of Γ.

   (This is known in the performance engineering literature as the "heavy traffic" theorem for queues.)

3. Suppose that V_0 is gamma distributed with shape parameter $\alpha = 2$ and scale parameter 1. What is an approximation to $P(W_\infty > 2)$ if $\lambda = 0.45$?

Extra Credit Problems

These extra credit problems do not serve as a replacement for the regular assigned problems. Rather, they are an "all or nothing" shot meaning you must get it entirely correct in order to get credit. At the end of the quarter, we will total the number you of ECP you have done and award an appropriate bonus.

Problem 1: Assume that Z is generated from r.v.'s X, Y (i.e., $Z = \varphi(X, Y)$ for some function φ). Suppose further that the times to generate X and Y are a and b, $a \gg b$, respectively. The simulator judges that only the Y taking values in a certain set E contribute significantly to $z = \mathbf{E}\varphi(X, Y)$ and wants to construct an approach that tries to cut down on unnecessary computational costs. He therefore does not necessarily generate Z if $Y \notin E$ but rather only if a coin toss comes out with heads; the probability of heads is p.

Write up an unbiased estimator for z according to the simulator's rules.

Problem 2:

1. Prove that if $Z = \tau_1 + \cdots + \tau_n$, where the τ_i's are iid Exp(λ) rvs, then Z has a Gamma($1/\lambda, n$) distribution.

2. Suppose the $1 \leq n < \alpha < n+1$. Use (a) to produce and acceptance-rejection algorithm for generating Gamma(λ, α) rvs.

Problem 3: Propose a kernel-based estimator for the density of an $\mathbb{R}^d-$ valued random variable X based on observing $X_1, X_2, \ldots, X_n$. Compute the optimal rate of convergence as a function of d.