MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2008
Problem Set 10 due 11/26/2008

Exercise 1. Let $\{X_n\}$ be a sequence of identically distributed random variables, with finite variance. Suppose that $\text{cov}(X_i, X_j) \leq \alpha^{|i-j|}$, for every i and j, where $|\alpha| < 1$. Show that the sample mean $(X_1 + \cdots + X_n)/n$ converges to $\mathbb{E}[X_1]$, in probability.

Exercise 2. Suppose that a random variable does not admit a finite upper bound, that is, $F_X(x) < 1$ for all $x \in \mathbb{R}$. Show that

$$\lim_{s\to\infty} \frac{\log M(s)}{s} = \infty.$$

Exercise 3. Let $X_1, X_2, \ldots$ be i.i.d. exponential random variables with parameter $\lambda = 1$. Let $S_n = X_1 + \cdots + X_n$. What is the Chernoff upper bound for $\mathbb{P}(S_n \geq na)$?

Exercise 4. ("Change of measure" for fast simulation.)
Consider a nonnegative random variable X whose PDF is close to being exponential, of the form

$$f_X(x) = g(x)e^{-x},$$

where $g(x)$ is a nonnegative function that satisfies $1/2 \leq g(x) \leq 2$ for all x, and $\int g(x)e^{-x}\,dx = 1$. Let a be a large constant. We wish to estimate $p = \mathbb{P}(X \geq a)$ using Monte Carlo simulation. We assume that we are able to generate random variables drawn from the distribution of X, as well as from an exponential distribution.

The straightforward simulation method is to generate n random samples, drawn from the distribution of X, let N be the number of samples that satisfy $X_i \geq a$, and form the estimate $\hat{P} = N/n$. Clearly, $\mathbb{E}[\hat{P}] = p$.

(a) Show that for large enough a, we have $\text{var}(\hat{P}) \geq e^{-a}/(3n)$.

Part (a) shows that the standard deviation of the estimation error $\hat{P} - p$ is of order $O(e^{-a/2})$, which is larger than the quantity p to be estimated by a $O(e^{a/2})$ factor. This is an instance of a general phenomenon: probabilities of rare events are difficult to estimate by simulation.

Consider now a random variable Y whose PDF is exponential, with parameter $\lambda = 1/a$.

$$f_Y(x) = \frac{e^{-x/a}}{a} = \exp\Big\{\Big(1 - \frac{1}{a}\Big)x\Big\} \cdot \frac{1}{a \cdot g(x)} \cdot f_X(x).$$

We generate n random samples Y_i, drawn from the distribution of Y, and estimate p by

$$Q = \frac{1}{n}\sum_{i=1}^{n} I_{Y_i \geq a}\frac{f_X(Y_i)}{f_Y(Y_i)} = \frac{1}{n}\sum_{i=1}^{n} I_{Y_i \geq a} \cdot ag(Y_i) \cdot \exp\Big\{-\Big(1 - \frac{1}{a}\Big)Y_i\Big\}.$$

(b) Show that $\mathbb{E}[Q] = p$.

(c) Show that the standard deviation σ_Q of Q is "comparable" to p, in the sense that σ_Q/p does not grow exponentially with a.

Exercise 5. A coin is tossed independently n times. The probability of heads at each toss is p. At each time k (with $k = 2, \ldots, n$), we obtain a unit reward at time $k + 1$ if the kth toss is heads and the previous toss was tails. Let R be the total reward obtained.

(a) Each time k (with $k < n$) that a tail is obtained, there is a probability p that the next toss is heads, in which case a unit reward is obtained at time $k + 1$. Let T be the number of tails in tosses $1, \ldots, n - 1$. Is it true that conditional on $T = t$, the reward R has a binomial (conditional) distribution with parameters t and p? Justify your answer.

(b) Let A_k be the event that a reward is obtained at time k.

   (i) Are the events A_k and A_{k+1} independent?

   (ii) Are the events A_k and A_{k+2} independent?

(c) Find the expected value of R.

(d) Find the variance of R.

(e) If the number n of coin tosses is infinite, what is the expected value of the number of tosses until the reward becomes equal to some given number k?

(f) Suppose that $n = 1000$ and $p = 1/2$. Find an approximation to the probability that $R \geq 260$. You may leave your answer in the form $\Phi(c)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable, and c is some number.

Exercise 6. For any positive integer k, let

$$h_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}.$$

Consider a Poisson process and let $X_k = 1$ if and only if there has been at least one arrival during the interval $[h_k, h_{k+1})$. Show that X_k converges to zero in probability, but not almost surely.

Exercise 7. Let $N(\cdot)$ be a Poisson process with rate λ. Find the covariance of $N(s)$ and $N(t)$.

Exercise 8. Based on your understanding of the Poisson process, determine the numerical values of a and b in the following expression and explain your reasoning.

$$\int_t^{\infty} \frac{\lambda^5 \tau^4 e^{-\lambda\tau}}{4!}\, d\tau = \sum_{k=a}^{b} \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Exercise 9. [Drill problem, does not have to be turned in.]
Fred is giving out samples of dog food. He makes calls door to door, but he leaves a sample (one can) only on those calls for which the door is answered *and* a dog is in residence. On any call the probability of the door being answered is $3/4$, and the probability that any household has a dog is $2/3$. Assume that the events "Door answered" and "A dog lives here" are independent and also that the outcomes of all calls are independent.

(a) Determine the probability that Fred gives away his first sample on his third call.

(b) Given that he has given away exactly four samples on his first eight calls, determine the conditional probability that Fred will give away his fifth sample on his eleventh call.

(c) Determine the probability that he gives away his second sample on his fifth call.

(d) Given that he did not give away his second sample on his second call, determine the conditional probability that he will leave his second sample on his fifth call.

(e) We will say that Fred "needs a new supply" immediately *after* the call on which he gives away his last can. If he starts out with two cans, determine the probability that he completes at least five calls before he needs a new supply.

(f) If he starts out with exactly m cans, determine the expected value and variance of D_m, the number of homes with dogs which he passes up (because of no answer) before he needs a new supply.

Hint: The formula $\text{var}(X) = \text{var}(\mathbb{E}[X \mid Y]) + \mathbb{E}[\text{var}(X \mid Y)]$ may be useful. Also, if T_i are i.i.d. geometric random variables, with parameter p, and $Y_k = T_1 + \cdots + T_n$, then the PMF of Y_k (known as a Pascal PMF) is of the form

$$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}, \qquad t = k, k+1, \ldots.$$

6.436J / 15.085J Fundamentals of Probability
Fall 2008