Thursday, December 11, 2003
Math 5040 Take-Home Final Exam with Solutions

Directions: You are to solve any 4 of the 5 problems. (There is no advantage to solving all 5.) Please proofread your work carefully before turning it in.

1. (Requires computing.) Let X be standard normal random variable. Evaluate $\theta = E[e^{-X^4}]$ by Monte Carlo simulation. (You can probably obtain a highly accurate numerical approximation to θ using Mathematica or Maple. That is not what we want.) By simulating the $N(0,1)$ distribution (see the book for several methods), you are required to estimate θ to within 0.001. First determine how many repetitions of the simulation you will need to achieve this degree of accuracy. Include any code and any output as part of your solution.

Solution: Let us use the algorithm on page 656 of the 8th edition. First we run a short program to estimate the variance.

```
RANDOMIZE
LET runs=500
FOR n=1 to runs
    LET S=2
    DO until S<=1
        LET V1=2*rnd-1
        LET V2=2*rnd-1
        LET S=V1^2+V2^2
    LOOP
    LET Z=(2*log(S)/S)^2
    LET X4=Z*V1^4  !  4th power of a standard normal
    LET Y4=Z*V2^4  !  4th power of a standard normal
    LET mean=mean+exp(-X4)+exp(-Y4)
    LET mom2=mom2+exp(-2*X4)+exp(-2*Y4)
NEXT n
PRINT mom2/(2*runs)-(mean/(2*runs))^2
END
```

We get a variance of about 0.16, hence a standard deviation of about 0.4, so the required sample size is about $(1.96*0.4/0.001)^2$, or about 615,000.

We take the sample size to be one million just to be safe. Each application of the algorithm produces two independent $N(0,1)$ random variables, so we need only a half million runs.

```
RANDOMIZE
LET runs=500000
FOR n=1 to runs
    LET S=2
    DO until S<=1
        LET V1=2*rnd-1
        LET V2=2*rnd-1
        LET S=V1^2+V2^2
    LOOP
    LET Z=(2*log(S)/S)^2
    LET X4=Z*V1^4  !  4th power of a standard normal
    LET Y4=Z*V2^4  !  4th power of a standard normal
    LET mean=mean+exp(-X4)+exp(-Y4)
NEXT n
PRINT mean/(2*runs) !  Print estimate of mean
END
```

output: 0.620533

2. Consider the Markov chain in the set of integers

$$S = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$$

with transition matrix P given by $P_{i,i+2} = p$, $P_{i,i-1} = q = 1 - p$, and $P_{ij} = 0$ otherwise. Under what conditions on p $(0 < p < 1)$ is the Markov chain recurrent? Hint: First find a formula for P_{00}^{3n}. You may need to use Stirling's formula to answer the question: $n! \sim n^n e^{-n} \sqrt{2\pi n}$. ("Example 4.15 (A Random Walk)" in the book is related to this.)

Solution: $P_{00}^{3n+1} = P_{00}^{3n+2} = 0$ and

$$\begin{aligned} P_{00}^{3n} &= \binom{3n}{n} p^n q^{2n} = \frac{(3n)!}{n!\,(2n)!}(pq^2)^n \\ &\sim \frac{(3n)^{3n}e^{-3n}\sqrt{6\pi n}}{n^n e^{-n}\sqrt{2\pi n}\,(2n)^{2n}e^{-2n}\sqrt{4\pi n}}\,(pq^2)^n \\ &= \frac{\left(\frac{27}{4}pq^2\right)^n}{\sqrt{\left(\frac{4}{3}\right)\pi n}}. \end{aligned}$$

Now $f(p) = \frac{27}{4}pq^2 = \frac{27}{4}p(1-p)^2$ is uniquely maximized at $p = 1/3$, where it is equal to 1. Therefore the Markov chain is recurrent iff $\sum P_{00}^{3n} = \infty$ iff

$$\sum_{n\geq 1} \frac{\left(\frac{27}{4}pq^2\right)^n}{\sqrt{\left(\frac{4}{3}\right)\pi n}} = \infty$$

iff $p = 1/3$.

3. Suppose $p_i > 0$ for $i = 1, 2, 3, \ldots$ and $\sum_{i=1}^{\infty} p_i = 1$. Consider the Markov chain in $S = \{0, 1, 2, \ldots\}$ with transition probabilities $P_{0j} = p_j$ for $j = 1, 2, 3, \ldots$; $P_{i0} = 1$ for $i = 1, 2, 3, \ldots$; and $P_{ij} = 0$ otherwise. Note that the Markov chain is irreducible.

(a) Find its period.

(b) Find the distribution of the first return time to state i (starting in state i). Treat the two cases $i = 0$ and $i \geq 1$ separately.

(c) Is the Markov chain transient, positive recurrent, or null recurrent?

(d) Is there a stationary distribution? If so, find it.

Solution: (a) $P_{00}^{2n+1} = 0$ and $P_{00}^{2n} = 1$, so the period of state 0 is g.c.d.$\{2, 4, 6, \ldots\} = 2$, and therefore this is the period of the Markov chain.

(b) $P_0\{T = 2\} = 1$, while for $i \geq 1$, $P_i\{T = 2n\} = (1 - p_i)^{n-1}p_i$, that is, the return time to state $i \geq 1$ is twice a geometric(p_i). In particular, the mean return time to state $i \geq 1$ is $m_{ii} = 2/p_i$.

(c) The mean return time to state 0 is 2, which is finite, so the Markov chain is positive recurrent.

(d) An irreducible positive recurrent Markov chain always has a stationary distribution, even if it is periodic (remark (iii), page 201). It is given by $\pi_i = 1/m_{ii}$ (remark (ii), page 208), so $\pi_0 = 1/2$ and for $i \geq 1$, $\pi_i = p_i/2$. Alternatively, we could use $\pi = \pi P$, that is, $\pi_0 = \pi_1 + \pi_2 + \cdots$ and $\pi_i = \pi_0 p_i$. Since $\pi_0 + \pi_1 + \pi_2 + \cdots = 1$, we find that $\pi_0 = 1/2$ and for $i \geq 1$, $\pi_i = p_i/2$.

4. Suppose that every man in a certain society fathers exactly two children, which independently have probabilities p of being a boy and $q = 1 - p$ of being a girl. Suppose also that the number of males in the nth generation, X_n, forms a branching process.

(a) Find the distribution of the number of boys in the second generation whose fathers are sons of a particular male in generation 0.

(b) Find the probability that the male line of a given man eventually becomes extinct. (The answer will depend on p.)

Solution: (a) The male offspring distribution from a male parent is 0 with probability q^2, 1 with probability $2pq$, and 2 with probability p^2. This is of course the binomial$(2, p)$ distribution. If there are no male offspring in generation 1, there will be none in generation 2. If there is one in generation 1, there will be a binomial$(2, p)$ number in generation 2. If there are two in generation 1, there will be a binomial$(4, p)$ number in generation 2. Putting all these results together, the number in generation 2 is 0 with probability $q^2 + 2pq^3 + p^2q^4$, 1 with probability $4p^2q^2 + 4p^3q^3$, 2 with probability $2p^3q + 6p^4q^2$, 3 with probability $4p^5q$, and 4 with probability p^6.

(b) The mean number of male offspring is the mean of binomial$(2, p)$, which is $2p$. If this is less than or equal to 1, then $\pi = 1$. If it is greater than 1, then $\pi < 1$ and π satisfies $q^2 + 2pq\pi + p^2\pi^2 = \pi$. Since 1 is always a root of this quadratic, we can factor it as $(\pi - 1)(p^2\pi - q^2) = 0$, hence $\pi = q^2/p^2$. To summarize, $\pi = 1$ if $p \leq 1/2$ and $\pi = (1/p - 1)^2$ if $p > 1/2$.

5. Let S be the set of all sequences of 0s and 1s of length 100 that have no adjacent 1s. (Without the restriction on adjacent 1s, there would be 2^{100} elements in S, but with the restriction, there is no known formula for the number of elements in S.) Use Markov Chain Monte Carlo methods to write an algorithm to simulate the expected number of 1s in an element chosen at random from S (i.e., uniformly distributed). Your algorithm should be specific enough that a programmer who knows nothing about stochastic processes could write the code directly from your description. Be sure to justify your method.

Solution: Consider the Markov chain in S that jumps as follows: Starting from any sequence in S, choose a position at random from $1, 2, \ldots, 100$ and switch a 1 to a 0, or a 0 to a 1 provided the new sequence belongs to S; if it doesn't, no change occurs. The transition matrix is clearly symmetric, so the unique stationary distribution is uniform. Therefore, we need only run the Markov chain for a long time (say a million steps), counting the number of 1s at each step, and average these results for our estimate.

A program is given below. It is convenient to have an array `seq` indexed by $0, 1, \ldots, 101$ with permanent 0s in positions 0 and 101. Then a change is made at state i if $\texttt{seq}(i) = 1$ or if $\texttt{seq}(i-1) = 0$ and $\texttt{seq}(i) = 0$ and $\texttt{seq}(i+1) = 0$; the latter condition can be written $(1 - \texttt{seq}(i-1))(1 - \texttt{seq}(i))(1 - \texttt{seq}(i+1)) = 1$.

```
DIM seq(0:101)
RANDOMIZE
LET steps=1000000
LET sum=0
FOR n=1 to steps
    LET i=int(100*rnd)+1
    IF seq(i)=1 or (1-seq(i-1))*(1-seq(i))*(1-seq(i+1))=1
        then LET seq(i)=1-seq(i)
    FOR j=1 to 100
        LET sum=sum+seq(i)
    NEXT j
NEXT n
PRINT sum/steps
END
```

output: 27.7966

Writing a program and getting a numerical estimate was not required, but it was very easy once the algorithm was precisely formulated.