Math 5920, Course Project Instructions:

**The task:** Your task is to use Monte Carlo methods, implemented in R, to solve some problem that is of interest to you. This problem can come from any field, and it need not be a problem that can *only* be solved using Monte Carlo methods. Your grade will be based on factors such as (1) the difficulty of the problem you solved, (2) the accuracy of your solution, (3) the quality and clarity of your R code, (4) the quality and clarity of your written explanation, and (5) the creativity that you showed in solving the problem.

**Type of work:** You may work either alone or as part of a team of size two. My expectations for teams are higher than my expectations for individuals.

**What to turn in:** Turn in (1) an annotated copy of your R code, (2) an electronic version of your R code, submitted via e-mail, (3) a written explanation of what your code does and what you’ve learned from your code, and (4) a list of the people and sources that you consulted. There are no restrictions on the length of the written explanation. It just needs to be long enough to explain what you did and what you found.

**Project topics:** Some potential project topics are listed on the next few pages. More than one individual or team can attempt the same topic, provided that the work is independent. However, if a topic proves to be too popular, I may ask you to choose a different one. You may also choose a project topic that is not on the list. If you have a potential topic in mind, I would be happy to help you develop it further.

**Due date for topic:** April 10, 2012. Please let me know your plan for the project by sending an e-mail. I also recommend scheduling a meeting with me to discuss your plan.

**Due date for project:** May 11, 2012. You are welcome to turn the project in earlier or to ask for feedback on your progress at any point.

**Outside consultation:** You are free to consult any person and any source. However, you must turn in a list of the people and sources that you consulted. Accessing good sources will be viewed as a point in your favor.
Math 5920, Some Potential Project Topics:

1. **Central Limit Theorem:** The Central Limit Theorem states that under appropriate conditions, averages of \( n \) random variables become more and more normally distributed as \( n \) increases. Using Monte Carlo methods, investigate how large \( n \) must be for the Central Limit Theorem to give a good approximation to the distribution of the sample mean. Your investigation should involve considering a wide variety of distributions and sample sizes. Summarize your results by constructing a few well-designed tables or figures.

2. **Robustness of the one-sample \( t \) test:** The one-sample \( t \) test is a test for whether the mean of a distribution equals some specified value or not, and the \( t \) confidence interval is a confidence interval for the population mean. Both methods require a normality assumption for validity. Using Monte Carlo methods, investigate how well the test and the confidence interval perform when this assumption is violated. Is it true that the test and the confidence interval are robust to departures from normality? Summarize your results by constructing a few well-designed tables or figures.

3. **Robustness of the two-sample \( t \) test:** The two-sample \( t \) test is a test for whether the means of two populations are equal or not. There are two versions of the test, one of which requires an assumption of equal variances and one of which does not. The test also requires a normality assumption for validity. Using Monte Carlo methods, investigate how well the two versions of the test perform when the normality and equal variances assumptions are violated. Is it true that the two-sample \( t \) test is robust to departures from normality? Summarize your results by constructing a few well-designed tables or figures.

4. **Testing normality:** Given a set of data, how do we decide whether the data are consistent with normality or not? Find out what tests are available, and compare those tests in terms of power using Monte Carlo methods. Your investigation should involve considering a wide variety of alternative distributions and sample sizes. Which tests would you recommend for use? Summarize your results by constructing a few well-designed tables or figures.

5. **Robustness of the \( F \) test for equal variances:** The \( F \) test is a test for comparing the variances for two distributions. Like the \( t \) test, it requires a normality assumption for validity. Using Monte Carlo methods, investigate how well the test performs when this assumption is violated. What alternative tests are available? How do the alternative tests compare to the \( F \) test in terms of maintaining the level of the test? How the alternatives compare to the \( F \) test in terms of power? Summarize your results by constructing a few well-designed tables or figures.
6. **Robustness of the ANOVA $F$ test:** The ANOVA $F$ test is a test for assessing whether the means of several different populations are equal or not. Like the $t$ test, it requires a normality assumption for validity. It also requires an assumption of equal variances. Using Monte Carlo methods, investigate how well the test performs when these assumptions are violated. Consider both the equal-sample-size case and the unequal-sample-size case. Summarize your results by constructing a few well-designed tables or figures.

7. **Coupling from the past:** Section 3.7.2 of the textbook discusses a Monte Carlo method called “coupling from the past”. Develop R code for implementing this method to sample from the stationary distribution of a Markov chain. Is coupling from the past more effective than simulating the Markov chain into the future? Can you implement coupling from the past for the “dots on a chess board” example that we discussed in class to motivate the need for studying Markov chains?

8. **Line-up simulation in baseball:** Given a particular list of nine baseball players, how should the players be ordered in order to maximize the expected number of runs scored in a game? Develop R code for simulating the performance of a given line-up, and use your code to compare different line-up choices for some selection of players. Do the results follow what you expected, or are there surprises? How much does line-up selection matter?

9. **Ranking teams using random walkers:** There are many different methods for ranking college football or basketball teams. Some of these methods involve Monte Carlo methods. One example is the “simple walker” model developed by Callaghan, Mucha, and Porter (2004). Develop code for implementing this ranking method, and consider developing a ranking method of your own that operates along similar lines. How do the methods differ? Which seems to be better? How does one decide which method is preferable?

10. **Redesigning the Big East basketball tournament:** Suppose that you have been selected to redesign the Big East basketball tournament. How does one design the tournament to maximize the chance that the true best team wins the tournament? Propose several different tournament designs, and compare their performance using Monte Carlo methods. Which design would you choose?

11. **Comparing measures of center:** There are a variety of different estimators for the center of a symmetric distribution. These include the sample mean, the sample median, trimmed means, and a variety of other choices. Using Monte Carlo methods, investigate how well these different estimators perform for estimating the center of the distribution. Your investigation should involve considering a wide variety of symmetric distributions and sample sizes. Summarize your results by constructing a few well-designed tables or figures.
12. **Simulating the Ising model:** Section 3.5 of the textbook discusses the use of Monte Carlo methods to simulate from the Ising model, which is used in physics. Develop code to simulate from the Ising model using different Metropolis-Hastings algorithms. Also try different dimensions and different energy levels. Do some versions of Metropolis-Hastings seem to work better than others?

13. **Robustness of Hartley’s test for equal variances:** Hartley’s test for equal variances is a test for comparing the variances of two or more distributions. Like the $t$ test, it requires a normality assumption for validity. Using Monte Carlo methods, investigate how well the test performs when this assumption are violated. Also develop code for finding critical values for the test using Monte Carlo methods. Summarize your results by constructing a few well-designed tables or figures.

14. **Comparing tests for uniformity:** There are many ways of testing whether data values $X_1, \ldots, X_n$ come from a uniform distribution on the interval $[0, 1]$. Pick out some of these tests and compare them in terms of power using Monte Carlo methods. Your investigation should involve considering a wide variety of alternative distributions and sample sizes. Also develop code for finding critical values for each test using Monte Carlo methods. Summarize your results by constructing a few well-designed tables or figures.