Exam **Mathematical Simulation**
Bachelor Knowledge Engineering, block 2.5
Thursday 3 June 2010, 09.00-12.00h, Bouillonstraat 8-10, room 0.015
Please make sure that your answers are to-the-point and concise, but ensure that you give a **proper explanation**! *Ensure that your cell phone is turned off and is not within reach!*
This is an open book exam. Please hand in the exam questions.

1. In order to secure a soccer match during the world cup, all visitors are frisk searched. There are 50 gates at which the visitors can be searched by a a pair of guards. If a metal object is found it is further investigated and classified as either harmless, forbidden or criminal. In the first case nothing happens, in the second case the item is confiscated and in the third case the visitor and item are turned over to the police, forcing the gate to close until the police has arrived.
   (a) Is this a terminating or steady state simulation? Why?
   (b) Give an event base description of this problem.

2. To cast a vote for the elections for the "Tweede Kamer der Staten Generaal" you first have to identify yourself by handing over your "stempas" and an ID. This takes an exponential distributed amount of time with rate 2 per minute and you get a ticket. With this ticket you have to get your polling card which takes again an exponential distributed amount of time, but then with rate 3 per minute. Finally you have to cast your vote by filling out your polling card in a booth taking an exponentially distributed amount of time with rate 2 per minute. When your arrive at the polling station you find that someone else is already in the polling booth. Other people in service have a blocking effect and there is ample queuing capacity.
   (a) Find the probability that the polling booth is still occupied when you have received your ticket.
   (b) Find the probability that the polling booth is still occupied when you receive your polling card.
   (c) Find the expected amount of time that you spend in the system.

3. Consider the following LCG with multiplier $a = 5$, increment $c = 3$, modulus $m = 8$ and seed $Z_0 = 7$.
   (a) Determine the period of this LCG.
   (b) Determine the value of Z_{333}.
   (c) Why can it be efficient for an LCG to take word length as modulus and what does this mean in terms of portability?

4. (a) Apply the acceptance-rejection method to generate a random variate with the density function
$$f(x) = \begin{cases} \frac{3}{2}x^2 & \text{if } -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$
   (b) Consider the following pdf:
$$f(x) = \begin{cases} \frac{2}{9\sqrt[3]{x}} & \text{if } 1 \leq x < \sqrt{8} \\ \frac{1}{3}x^3 & \text{if } \sqrt{8} \leq x < \sqrt[4]{72} \\ 0 & \text{otherwise} \end{cases}$$
   Give the *inverse-transform* algorithm for generating random variates from this distribution.

5. For a poker test that is performed, as discussed during the lecture, on non-overlapping sequences of five consecutive integers $\lfloor u_k \rfloor$, the probability that such a sequence is of a certain category is given by:

Category	Probability
all different	0.3024
one pair	0.5040
two pairs	0.1080
three of a kind	0.0720
full house (three of a kind and a pair)	0.0090
four of a kind	0.0045
five of a kind	0.0001

Eddy determines the number of occurrences in each category for a certain data set and these are his findings:

Category	Number of occurences
all different	25
one pair	53
two pairs	12
three of a kind	7
full house (three of a kind and a pair)	2
four of a kind	0
five of a kind	1

Perform a poker test on the provided data to check the hypothesis that the data is random.

6. A Dutch contractor has offered to help clean up the oil in the Gulf of Mexico and he sends in the following independent test data of liters cleaned up oil per ship per hour: 2699, 2778, 3037, 902, 5285, 4172, 3092, 5217, 2135, 3744, 3605, 4166, 4159, 2982, 3816, 3682, 4474, 4940, 4956, 2983. From experience he knows that this data is normally distributed.

   The oil company prefers an American contractor and asks a rivalling contractor for similar test data as well yielding: 4151, 3211, 4297, 2638, 2007, 1981, 3891, 3225, 3206, 4823, 3694, 3600, 4990, 2598, 4099, 4238, 3159, 3618, 2237, 2255

   (a) The secretary of the oil company's CEO says that she has determined the average of oil cleanup for each company and has determined that the Dutch company is the best one. The CEO's son who happens to be a knowledge engineer says that this is not the proper way of drawing conclusions from this data. What arguments could he give here?

   (b) Can you determine using a 90% confidence interval on the difference determine which company is the best?

   (c) The CEO want to be quite sure on who he should hire. He doesn't care about 100 liters more or less, so as long as the difference is smaller than that he doesn't care about who he hires, but if the difference is larger he wants the be 95% sure that he has hired the best company. How many additional test samples should each of the companies provide to determine the best one using the algorithm of *Dudewicz and Dalal*, taking into account that instead of a minimization problem we deal with a maximization problem?

   (d) Suppose that the additional tests as required under (c) are performed and the resulting second-stage mean cleanup volumes are $\overline{X}^{(2)}_{\text{Dutch}} = 3847$ and $\overline{X}^{(2)}_{\text{USA}} = 3403$. Can the CEO now select the *fastest cleaning company* and, if "yes", what company should he select?

7. Consider the following Monte Carlo example where U represents a random number:

$$X_j = (1 - U^2)^2$$

   Analytically determine $\text{Cov}(X_j^{(1)}, X_j^{(2)})$ and analytically calculate $\text{Var}(\frac{X_j^{(1)} - X_j^{(2)}}{2})$ under both independent sampling and antithetic variables. What are your conclusions?

Success!