NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for
TMA4300 Computer Intensive Statistical Methods

Academic contact during examination: Geir-Arne Fuglstad

Phone: 4527 0806

Examination date: May 19th 2015

Examination time (from–to): 09:00–13:00

Permitted examination support material: C:

- Calculator HP30S, CITIZEN SR-270X or CITIZEN SR-270X College, Casio fx-82ES PLUS with empty memory.
- Tabeller og formler i statistikk, Tapir.
- K. Rottmann: Matematisk formelsamling.
- One yellow, stamped A5 sheet with own handwritten formulas and notes.
- Dictionary in any language.

Other information:

- All seven sub-problems in this exam count approximately the same.
- All answers must be justified.
- In your solution you can use English and/or Norwegian.

Language: English

Number of pages: 5

Number pages enclosed: 0

Checked by:

Date Signature

Problem 1

a)
- Describe shortly two applications for stochastic simulation.
- What are pseudo-random numbers?

Problem 2

The 2D-vectors shown in the following figure are labelled using either squares or circles. Use the k-nearest-neighbour approach to classify the point $(6,6)$, which is indicated by a triangle in the figure. As a distance function use the L_1-norm which is for two 2-dimensional vectors $\mathbf{p} = (p_1, p_2)$ and $\mathbf{q} = (q_1, q_2)$ defined as:

$$d(\mathbf{p}, \mathbf{q}) = \mid p_1 - q_1 \mid + \mid p_2 - q_2 \mid$$

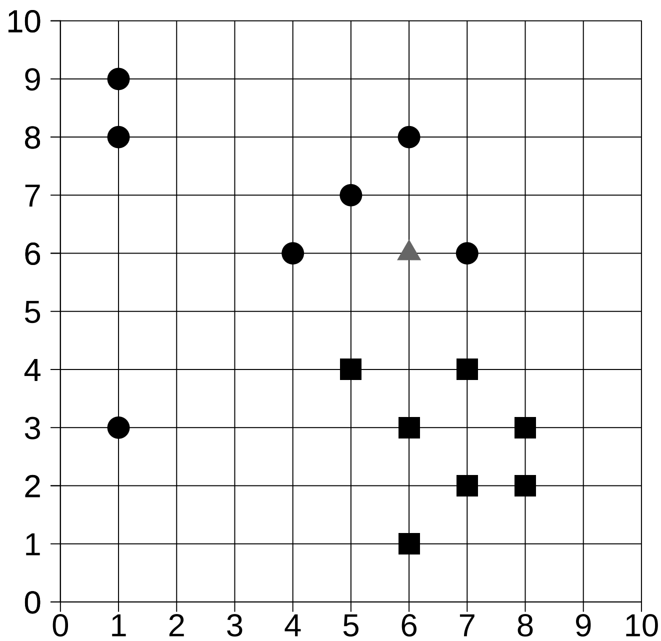
i.e. the sum of the absolute differences of their Cartesian coordinates.

a) More specifically, classify the point $(6,6)$ using the following values of k:

- $k = 4$
- $k = 7$
- $k = 10$

Considering not only this example, what can be a problem when k is chosen too small or too big?

Which procedure would you propose for choosing a sensible value of k? Describe the approach.



Problem 3

a) Propose a Metropolis-Hastings algorithm to sample according to a binomial distribution with density

$$p(x) = \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1-p)^{n-x}, \quad x = 0, 1, 2, 3, \ldots, n,$$

where n denotes the number of trials and p the success probability. Assume thereby that you cannot sample from the binomial distribution directly. Describe all steps of the algorithm and be as precise as possible in your description (i.e. simplify all expressions etc.).

Problem 4

Assume we have a data set of n distinct observations.

a)
- Explain shortly how you obtain a bootstrap sample from the observed data set.
- Show that the probability that a certain observation appears in the bootstrap sample is ≈ 0.632 as $n \to \infty$.

b) Among all possible bootstrap samples (ignoring order), what is the most likely bootstrap sample? Derive the probability to obtain this bootstrap sample.

Problem 5

a) Let $Y \sim \text{Exp}(\lambda)$. Further, let $X = Y|Y > a$ where $a > 0$ is fixed. That means, the random variable X is equal to Y conditioned on $Y > a$. Find a formula for the cumulative distribution function $F_X(x)$ and the inverse of $F_X(x)$. Give an algorithm for simulating X using $U \sim \text{Unif}(0, 1)$.

b) An alternative algorithm to simulate $X = Y|Y > a$ for $Y \sim \text{Exp}(\lambda)$ is

1. Simulate $Y \sim \text{Exp}(\lambda)$
2. If $Y > a$, then stop and return $X = Y$, otherwise go back to 1.

Explain why this algorithm is just a rejection-sampling algorithm. To do this, define the target density f and the proposal density g. Derive also the bound c, so that $c = \max_x f(x)/g(x)$. Show that the rejection-sampling algorithm reduces to the algorithm described above in steps 1 and 2, and illustrate your explanation with a figure.

For $\lambda = 1$ and $a = 4$ what is the expected number of trials up to the first accepted sample?