Solution Sketch (not all details given)
Exam FS2015 TMA4300

05/10/15

1 Problem 1

-
  - Computing analytical results from complicated system, e.g. within physics, is often difficult. Simulation can be used to verify plausibility arguments.
  - Approximation of integrals/expected values
  - Simulation of stochastic processes for example within financial mathematics.
  - etc

  It is not sufficient to mention an application. It needs to be described shortly (more than done above), so that it becomes clear where/why simulation is involved and needed.

- Pseudo-random numbers are numbers in $[0, 1]$ that were generated using a deterministic function and have the same relevant statistical properties as a sequence of Unif$(0, 1)$ numbers. Given a particular function and a "seed" value, the same sequence of numbers will be generated by the function.

2 Problem 2

For $k = 4$ the point is classified as a circle, for $k = 7$ as well and for $k = 10$ it is classified as a square. This classification should be justified.

The problem with a too small chosen k is that the algorithm is sensitive to outliers, while when k is chosen too large many objects from other classes can be in the decision set. Figure 1 illustrates this.

M-fold cross validation could be used to select k (Here, the crucial details of the algorithm should be given). The algorithm is applied for each potential k and k is then chosen according to the smallest misclassification rate.

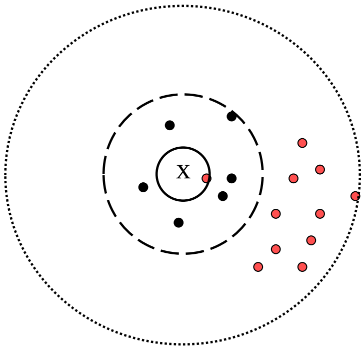

Figure 1: The object to be classified is indicated by a `x`. The solid circle indicates the decision set using $k = 1$, the dashed circle the set for $k = 7$ and the dotted circle the set for $k = 17$.

3 Problem 3

Proposing and describing a Metropolis-Hasting algorithm for example using a discrete uniform distribution on $\{0, \ldots, n\}$ as a proposal distribution or a random walk proposal. Using a random walk proposal be careful about boundaries. Continuous propoasal distribution are not sensible, also problem with boundaries appears. All steps (initialisation, i.e integer!,proposal distribution, acceptance rates, acceptance step, . . .) should be given in detail. Does something cancel?. Be careful that algoritm should run until convergence otherwise the generated samples will not be from the target binomial distribution. Log-scale should be preferred to compute acceptance ratio due to numerical problems in factorial computation.

It is insufficient to state the general Metropolis-Hastings algorithm, it needs to be adapted to the binomial target distribution, where all steps are specified and written out.

4 Problem 4

a)
- A bootstrap sample can be obtained by drawing n times with replacement from the observations of the original dataset. Each observation has the probability $1/n$ to be drawn. The resulting set has the same size as the original dataset, but some observations might be missing while others appear several times.

- The probability that a certain observation is not picked when we draw once is $1 - \frac{1}{n}$. Drawing n times the entry will not be picked with probability $\left(1 - \frac{1}{n}\right)^n$. Letting n go to infinity we get:

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} = 0.368.$$

  Thus, the probability that the observation appears in the bootstrap sample is approximated by $1 - 0.368 = 0.632$.

b) We regard the bootstrap sample as a set of observations, i.e. ordering is ignored. Since all entries are distinct, we can regard them as n different classes. Generating a bootstrap sample can be regarded as sampling from an urn with n different coloured balls with replacement. This is equivalent to sampling from a multinomial distribution. Thus each bootstrap sample has probability:

$$\frac{n!}{x_1! \cdots x_n!} \cdot \left(\frac{1}{n}\right)^{x_1+\cdots+x_n} = \frac{n!}{x_1! \cdots x_n!} \cdot \left(\frac{1}{n}\right)^{n}$$

Here, x_i represents the number of times the i-th observation is picked. The probability gets largest, namely $n!/n^n$, if $x_1 = \ldots = x_n = 1$ that means each entry appears once, i.e. we get the original data set.

5 Problem 5

a) $Y \sim \text{Exp}(\lambda)$. $X = Y \mid Y > a$ with $a > 0$.

$$\begin{aligned} F_X(x) = P(X \leq x) &= P(Y \leq x \mid Y > a) \\ &= \frac{P(Y \leq x, Y > a)}{P(Y > a)} \\ &= \frac{P(Y \leq x, Y > a)}{1 - P(Y \leq a)} \end{aligned}$$

If $x \leq a$ we get $P(Y \leq x, Y > a) = 0$ and thus $F_X(x) = 0$.
If $x > a$ we get:

$$\begin{aligned} P(Y \leq x, Y > a) &= P(Y \leq x) - P(Y \leq a) \\ &= 1 - \exp(-\lambda x) - (1 - \exp(-\lambda a)) = \exp(-\lambda a) - \exp(-\lambda x). \end{aligned}$$

Thus,

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{\exp(-\lambda a) - \exp(-\lambda x)}{\exp(-\lambda a)} = 1 - \frac{\exp(-\lambda x)}{\exp(-\lambda a)} & x > a. \end{cases}$$

This corresponds to the CDF of a shifted exponential distribution. The inverse CDF results as

$$F_X^{-1}(u) = a - \frac{1}{\lambda} \log(1 - u).$$

The inversion method can be used to generate samples for X. For this sample $U \sim \text{Unif}(0, 1)$ and evaluate $F_X^{-1}(u)$.

b) From a) we get

$$f_X(x) = \begin{cases} 0 & x \leq a \\ \frac{\lambda}{\exp(-\lambda a)} \exp(-\lambda x) & x > a. \end{cases}$$

This corresponds to our target density. The proposal density is the exponential distribution with parameter λ:

$$g_X(x) = \begin{cases} 0 & x < 0 \\ \lambda \exp(-\lambda x) & x \geq 0. \end{cases}$$

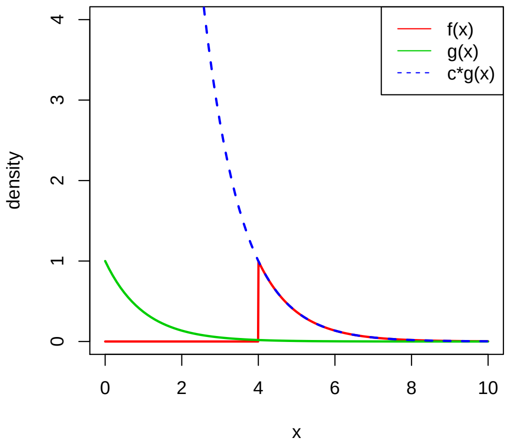

Figure 2: Rejection sampling setting for $\lambda = 1$ and $a = 4$.

From this we get:

$$\frac{f_X(x)}{g_X(x)} = \frac{\frac{\lambda}{\exp(-\lambda a)} \exp(-\lambda x)}{\lambda \exp(-\lambda x)} = \frac{1}{\exp(-\lambda a)} = \exp(\lambda a).$$

for $x > a$. That means $c = \exp(\lambda a)$. For $\lambda = 1$ and $a = 4$ we get $c = \exp(4) \approx 54$, which corresponds to the expected number of trials up to the first accepted sample. This is because we asssume that the trials are independent, so the number of trials up to the first success is geometrically distributed with parameter $1/c$.

Figure 2 illustrates $f(x)$, $g(x)$ and $c * g(x)$ for $\lambda = 1$ and $a = 4$. Since $f(x) = 0$ for all $x \leq a$ will all proposed samples from $g(x)$ that are smaller than a be rejected as $f(x)/(c \cdot g(x)) = 0$. On the other side all proposed samples that are larger than a will be accepted as $f(x)/(c \cdot g(x)) = 1$. Hence, it is enough to sample from $g(x)$, i.e. $X \sim \mathrm{Exp}(\lambda)$ and accept the value if it is larger than a.