Graz University of Technology

Department of Mathematical Structure Theory

Diskrete Stochastik und Informationstheorie – 21 May 2014

Exercise 27. The Google PageRank algorithm uses the stationary distribution of a network graph to estimate the relevance of web pages for its search results. The network graph represents the expected click behaviour of a user and is constructed as follows (simplified):

- Network graph: Each website is represented by a graph node. If website A links to website B (no matter if with just one or multiple links), then the network graph has a directed edge from node A to node B. The weights of all outgoing edges of node A are equal and sum up to 1.
- Transition graph: All edge weights from the network graph are multiplied by 0.85 ("damping factor"). Then, additional edges are inserted to construct a complete directed graph where each node is connected to each other node (but not to itself). The missing 0.15 (or 1 for nodes without outgoing edges in the network graph) to complete the outgoing transition probability distribution for each node are then distributed uniformly among outgoing edges of a node (both old and new edges).

The final transition graph corresponds to a stationary Markov chain. The rank of each website is then computed by finding (or approximating) the stationary distribution of this process (and the website with the highest probability is ranked highest).

(a) Pick a small imaginary network (at least 4 websites; for example: A links to B, C, and D; B to C and D; C to D; D has no links). Draw the corresponding network and transition graph with its transition probabilities.

(b) Compute the graph's stationary distribution. (Why) is it unique? (You can use a computer for the eigenvector computations. You can give either exact results or an approximate distribution starting from a uniform initial distribution.)

(c) What is the asymptotic entropy of this transition graph?

Exercise 28. (Flipping walker) A particle walks on the integers in the direction he is facing (left or right), reversing direction after each step taken with probability $p = 0.2$. The walker starts at 0 facing to the right.

(a) Describe the state space and transition probabilities of this process. Is (X_n) a Markov chain? Why/why not?

(b) What is the expected number of steps taken by the walker before reversing direction? *Hint:* Evaluate the series as in the hint of Exercise 15.

(c) Find $H(X_1, \ldots, X_n)$. *Hint:* The chain rule might be useful.

(d) Find the asymptotic entropy of this walker.

Exercise 29. For a stationary stochastic process, show the following properties:

(a) $\frac{1}{n}H(X_1, \ldots, X_n) \leq \frac{1}{n-1}H(X_1, \ldots, X_{n-1})$, and

(b) $\frac{1}{n}H(X_1, \ldots, X_n) \geq H(X_n \mid X_{n-1}, \ldots, X_1)$.

Hint: Useful ingredients (besides the definition of a stationary process) include

- the chain rule, $H(X_n, \ldots, X_1) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1)$,
- its special case $H(X_n, \ldots, X_1) = H(X_{n-1}, \ldots, X_1) + H(X_n \mid X_{n-1}, \ldots, X_1)$, and
- a corollary of the information inequality: $H(X \mid Y) \leq H(X)$.