Aalto University School of Electrical Engineering
Department of Communications and Networking

Exam 21.10.2015
Hyytiä & Peuhkuri

ELEC-E7130 - Internet Traffic Measurements and Analysis
S-38.3184 - Network Traffic Measurements and Analysis

1. **Explain** the following terms and **discuss** their roles in network measurements: (6 p)
   a) Scatterplot
   b) QQ-plot
   c) Time plot
   d) Association rule
   e) Outlier
   f) Classification (in data mining)

2. Distributions and fitting: Pareto distribution is defined by two parameters, shape α and scale k. Its mean and variance are

$$\mathrm{E}[X] = \frac{\alpha k}{\alpha - 1}, \qquad \sigma^2 = \left(\frac{k}{\alpha - 1}\right)^2 \cdot \frac{\alpha}{\alpha - 2}.$$

   (You can assume that $\alpha > 2$ so that both the mean and variance are finite)

   a) What is the second moment $\mathrm{E}[X^2]$ of the Pareto distribution? (2 p)
   b) IT department has collected data on the file sizes at the company's server and obtained $S = \{2, 2, 2, 6\}$ kB (i.e., sample size is 4). Fit the Pareto distribution to this data set using *the method of moments*. (3 p)
   c) What is the empirical cumulative distribution function of S? (1 p)

Following answers to a separate paper.

3. Define flow. Why a flow is important concept in network measurements? What roles granularity and timeout have with flows? (6 p)

4. How one can characterize network delay distributions? How these can be measured? What are the most important factors to consider when determining accuracy and errors of measurements? (6 p)

5. You work as a network engineer for a network operator. Your manager have been tasked to find out out how your customers are using Internet to better optimize peering agreements and possibly to provide some add-on services to customers. She hands out the task for you and expects a plan by tomorrow.

   Give a short description of setup and analysis. How you take care of user privacy in handling real, pseudonymous and anonymous identifiers? (6 p)