Institut EURECOM 13/04/2007
SSP Statistical Signal Processing

Homework 1

Due: 11/05/2007 (HW1 to be turned in to Antony Schutz.)

Homework policy: students are encouraged to discuss with fellow students to try to find the main structure of the solution for a problem, especially if they are totally stuck at the beginning of the problem. However, they should work out the details themselves and write down in their own words only what they understand themselves.

Problem: ML Estimation of Roundtrip Delay Distribution

In this problem, we consider the roundtrip delay in a computer network (internet) between the computer we're working on and another computer connected to the network. This roundtrip delay will be different, every time we send a message. As such, it can be modeled as a random variable y. For the design of network protocols and for their performance evaluation, it is important to know the distribution of this random roundtrip delay. To turn the estimation of the roundtrip delay distribution into a parameter estimation problem, we shall take a parametric distribution, parameterized by one or more parameters. Since we don't have too much information about this variable y (except that it should be positive), we shall try several parametric distributions. In particular, we shall consider the following distributions:

- A Gaussian distribution: $f_G\left(y|\mu,\sigma^2\right) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(y-\mu)^2}{2\sigma^2}}$
- A Rayleigh distribution: $f_R\left(y|\sigma^2\right) = \begin{cases} 0 & , \ y < 0 \\ \dfrac{y}{\sigma^2} e^{-\dfrac{y^2}{2\sigma^2}} & , \ y \geq 0 \end{cases}$
- An Erlang distribution: $f_{Em}\left(y|\lambda\right) = \begin{cases} 0 & , \ y < 0 \\ \dfrac{\lambda^{m+1}}{m!} y^m e^{-\lambda y} & , \ y \geq 0 \end{cases}$ for different values of $m \in \{0, 1, 2\}$.
  Remark that for $m = 0$ we have an exponential density: $f_{E0}\left(y|\lambda\right) = \begin{cases} 0 & , \ y < 0 \\ \lambda e^{-\lambda y} & , \ y \geq 0 \end{cases}$
- A (shifted) exponential density: $f_{exp}\left(y|\lambda,\alpha\right) = \begin{cases} 0 & , \ y < \alpha \\ \lambda e^{-\lambda(y-\alpha)} & , \ y \geq \alpha \end{cases}$

(i) Assume we collect n i.i.d. measurements $y_1, \ldots, y_n$ that we can put into a vector $Y = [y_1, y_2, \ldots, y_n]^T$. For each of the parametric distributions $f_i(Y|\theta)$, $i \in \{G, R, E0, E1, E2, exp\}$, determine the Maximum Likelihood estimate $\widehat{\theta}_{ML,i}$ of the parameter(s) θ involved.

(ii) Consider now also a shifted Rayleigh distribution:

$$f_{SR}(y|\alpha,\sigma^2) = \left\{ \begin{array}{ll} 0 & , y < \alpha \\ \dfrac{y-\alpha}{\sigma^2}\, e^{-\dfrac{(y-\alpha)^2}{2\sigma^2}} & , \alpha \le y \end{array} \right\} = \frac{y-\alpha}{\sigma^2}\, e^{-\dfrac{(y-\alpha)^2}{2\sigma^2}}\, \mathbf{1}_{[\alpha,\infty)}(y)$$

for some $\alpha \ge 0$, $\sigma^2 > 0$, and where we introduced the indicator function for a set A :

$$\mathbf{1}_A(y) = \left\{ \begin{array}{ll} 0 & , y \notin A \\ 1 & , y \in A \end{array} \right.$$

(a) Determine the mean of y, m_y, according to this shifted Rayleigh distribution, as a function of the parameters α and σ^2.

(b) Determine the variance of y, σ_y^2, according to this shifted Rayleigh distribution, as a function of the parameters α and σ^2.

(c) We now collect n i.i.d. measurements y_i into the vector Y.
Find the log likelihood function $L(\alpha, \sigma^2; Y)$ for α and σ^2 given Y.

(d) Reduce the range of possible values for $\alpha \ge 0$ by determining the range of α for which the log likelihood takes on finite values ($> -\infty$). Assume for what follows that α is in this range.

(e) For a given α, find the $\widehat{\sigma^2}(Y,\alpha)$ that maximizes the log likelihood function.

(f) Find the $\widehat{\alpha}_{ML}(Y)$ that maximizes $L(\alpha, \widehat{\sigma^2}(Y,\alpha); Y)$.

Determine the corresponding $\widehat{\sigma^2}_{ML}(Y) = \widehat{\sigma^2}(Y, \widehat{\alpha}_{ML}(Y))$.

(iii) In Matlab, you will generate $n = 100$ measurements of the roundtrip delay between your machine and a machine of your choice (preferably a machine that is not too close by). The roundtrip delay can be measured using the Unix command ping. The generation of 100 measurements, put into a vector that can be processed by Matlab, can be done by running the following command in Matlab

Y=pingstats('machine',100,'v');

for which the file pingstats.m has been distributed by email (put pingstats.m in the directory in which you launch Matlab; the comments in the file pingstats.m contain a suggestion for a machine to ping, but you are encouraged to try other machines, see the file hw_host_ip_list.txt which was also distributed by email). The file pingstats.m (and the command pingstats) is valid on Unix machines. On Linux machines, please use the file pingstats_forLinux.m (and the command pingstats_forLinux).

So, you generate a vector of 100 i.i.d. measurements for the roundtrip delay to one particular machine. In Matlab, calculate $\widehat{\theta}_{ML,i}$ using the expressions you derived previously, and this for each of the distributions $i \in \{G, R, E0, E1, E2, exp, SR\}$.

(iv) In Matlab, plot a histogram of the measurements $\{y_1, \ldots, y_n\}$ you made and in the same plot superimpose the graphs for the marginal densities $f_i(y|\widehat{\theta}_{ML,i}(Y))$, $i \in \{G, R, E0, E1, E2, exp, SR\}$. Make this plot for y going from a value that is somewhat smaller than the y_{min} you measured to a value that is somewhat bigger than the y_{max} you measured.

(v) Exploiting the Maximum Likelihood criterion to the fullest, we shall determine the best choice for the distribution of the roundtrip delays as the one that maximizes the likelihood:

$$\widehat{\imath}_{ML} = \arg \max_{i \in \{G,R,E0,E1,E2,exp,SR\}} f_i(Y|\widehat{\theta}_{ML,i}(Y)) \ .$$

In Matlab, calculate which is the best distribution in your case. Is the result in agreement with the graphs of the densities and the histogram you plotted?