Additional info Project 1 & 2: Handling the estimated uncertainty

SSY 230, System Identification
Jonas Sjöberg
Electrical Engineering, Chalmers

March 15, 2018

This document describes how you can use the estimated parameter uncertainty to describe the uncertainty of functions depending on the parameters. This is useful in project 1 and 2.

Given N data, the uncertainty of the obtained estimate $\hat{\theta}_N$ is described by the estimate of its variance P_θ. How P_θ is estimated is given by, eg, in S & S (4.12, 4,13) or R.J. (5.24).

The parameter estimate can the be considered as a realization from the following multi-dimensional normal distribution

$$\frac{1}{\sqrt{(2\pi)^2 \det P_\theta}} \exp(-\frac{1}{2}(\hat{\theta}_N - \theta_0) P_\theta^{-1} (\hat{\theta}_N - \theta_0)^T) \tag{1}$$

From this follows that

$$(\hat{\theta}_N - \theta_0)^T P_\theta^{-1} (\hat{\theta}_N - \theta_0)) \in \chi^2(d) \tag{2}$$

where

$$d = \dim \theta$$

Strictly, the distributions given above are only correct in the case that N goes to infinity, but the result hold approximately for finite, large, N. Therefor we say that we have an *asymptotic* normal distribution, and an *asymptotic* χ^2 distribution.

The total volume of the distribution function (1) is 1, ie,

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^2 \det P_\theta}} \exp(-\frac{1}{2}(\hat{\theta}_N - \theta_0)^T P_\theta^{-1} (\hat{\theta}_N - \theta_0)) d\theta_1 \dots d\theta_d = 1 \tag{3}$$

This integral is over the entire θ space. A confidence region, eg, 95% confidence region, is an ellipsoid so that the integral over the ellipsoid becomes 0.95. We are interested to generate number of examples, say n examples, on the boarder of this ellipsoid as examples of the uncertainty of the parameter estimate. It is, in fact, very easy to generate such examples, and that can be done by following this algorithm.

1. For each $k = 1$ to n.

2. Generate a random vector $\Delta\theta_k$ of dimension d and scale it so that its length becomes $\parallel \Delta\theta_k \parallel^2 = \chi^2_\alpha(d)$, where α is the confidence level you want to illustrate.

3. Generate the kth example vector as

$$\theta_k = \hat{\theta} + P^{1/2}\Delta\theta_k$$

Now, assume you have a function $f(x, \theta)$ and you have estimate $\hat{\theta}_N$ and $\{\theta_k\}_{k=1}^n$ obtained as described above. The function $f(x, \theta)$ can be an estimate of a static function, as in project 1, or a model of a dynamic system which depends on the parameters in a complicated way, as in a prediction in project 2. In any case, a function describing the upper limit of the uncertainty can be formulated as

$$f_{\max}(x, \hat{\theta}) = \max_k f(x, \theta_k) \tag{4}$$

A function for the lower limit, $f_{\min}$, can be defined similarly.

How good are these estimated of upper and lower limits? Well, they depend on the quality of the estimate of P_θ, N, and also how the data were generated, ie, how close to the assumptions they were generated. The quality also depends on how many random examples, n, you generated. You can try, eg, 5, 10 and 20 to see how much that influence the result.