Estimating the trajectory of a free falling object

The vertical coordinate ("height") of an object in free fall is described by an equation of the form

$$x(t) = \theta_0 + \theta_1 t + \theta_2 t^2,$$

where θ_0, θ_1, θ_2 are unknown parameters and t stands for time.[1] At certain times $t_1, \ldots, t_n$, we make noisy observations $y_1, \ldots, y_n$ of the height of the object. Based on these observations, we would like to estimate the object's true trajectory.

Let us cast this problem as a Bayesian inference problem. We view θ_0, θ_1, θ_2 as the realized values of continuous random variables Θ_0, Θ_1, Θ_2, with a given prior joint distribution. For simplicity, we assume that these random variables are independent, and denote the prior of Θ_i by $f_{\Theta_i}(\cdot)$. Our observation model is of the form

$$Y_i = X_i + W_i, \qquad i = 1, \ldots, n,$$

where

$$X_i = \Theta_0 + \Theta_1 t_i + \Theta_2 t_i^2$$

is the random height at the time t_i of the ith observation, and where W_i models the observation noise. We assume that the W_i are i.i.d., with a (common) PDF $f_W(\cdot)$, and that they are also independent from the Θ_i.

At this point, we have enough modeling assumptions in place to determine any PDF (marginal, joint, or conditional) of interest. Let us write down the form of the conditional (joint) PDF of $Y = (Y_1, \ldots, Y_n)$, given $\Theta = (\Theta_0, \Theta_1, \Theta_2)$. Once a value θ of Θ is specified, the values of the X_i are fixed to $x_i = \theta_0 + \theta_1 t_i + \theta_2 t_i^2$. Thus, in this conditional universe, Y_i is equal to $\theta_0 + \theta_1 t_i + \theta_2 t_i^2 + W_i$. Hence,

$$f_{Y_i|\Theta}(y_i \mid \theta) = f_W(y_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2).$$

Since the W_i are independent, it follows that the Y_i are conditionally (given Θ) independent. Thus,

$$f_{Y|\Theta}(y \mid \theta) = \prod_{i=1}^{n} f_{Y_i|\Theta}(y_i \mid \theta) = \prod_{i=1}^{n} f_W(y_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2).$$

[1] If the object is only subject to gravitational forces, the constant θ_2 would be known, fully determined by the gravitational constant g. However, if other effects are present, θ_2 may not be accurately known.

We can now determine the joint PDF of Θ and Y:

$$f_{\Theta,Y}(\theta, y) = \prod_{j=1}^{3} f_{\Theta_j}(\theta_j) \cdot \prod_{i=1}^{n} f_W(y_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2).$$

The posterior PDF is of the form

$$f_{\Theta|Y}(\theta \mid y) = \frac{1}{f_Y(y)} f_{\Theta,Y}(\theta, y),$$

where

$$f_Y(y) = \int f_{\Theta,Y}(y, \theta)\, d\theta.$$

Having observed the vector y of observations, we can evaluate numerically the posterior for any value of θ and plot it as a function of θ. As a practical matter, the hardest part of generating such a plot is the numerical evaluation of the three-dimensional integral $f_Y(y) = \int f_{\Theta,Y}(y, \theta)\, d\theta$. On the other hand, if we are interested in just the MAP estimator, or the general shape of the posterior distribution, the numerical value of $f_Y(y)$ is not required.

MAP estimation, assuming normal distributions

The problem is considerably simplified if we assume that the random variables $\Theta_0, \Theta_1, \Theta_2, W_1, \ldots, W_n$ are normal (and, as before, independent). For simplicity, let us assume that they all have zero-mean, that the Θ_i have variance σ_i^2, and that the W_i have a variance of σ^2. It follows that the (joint) posterior PDF is of the form

$$c(y) \exp\Big\{ -\frac{1}{2}(\theta_0^2/\sigma_0^2 + \theta_1^2/\sigma_1^2 + \theta_2^2/\sigma_2^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2 \Big\},$$

where $c(y)$ is a normalizing constant.

The MAP estimate is found by maximizing the posterior PDF: while keeping $y = (y_1, \ldots, y_n)$ fixed to the observed observation vector, minimize, with respect to $\theta = (\theta_0, \theta_1, \theta_2)$, the expression

$$\theta_0^2/\sigma_0^2 + \theta_1^2/\sigma_1^2 + \theta_2^2/\sigma_2^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 t_i - \theta_2 t_i^2)^2.$$

We are dealing with a quadratic function of θ; in particular, the partial derivatives (with respect to each θ_i) are linear functions of θ. To minimize this function, we set its three partial derivatives to zero and solve the resulting system of three linear equations in three unknowns.[2] Even though an explicit form of the solution might be a bit cumbersome to write down, the numerical calculation of the MAP estimate is very simple.

Remarks:

(a) Quadratic functions (of multiple variables) that have a unique minimum are always symmetric around their minimum. One can use this property to argue that the MAP estimate, $(\hat{\Theta}_1, \hat{\Theta}_2, \hat{\Theta}_3)_{\mathrm{MAP}}$, is the same as the vector with coordinates $\mathbf{E}[\Theta_i \mid Y]$.

(b) It can be verified that the conditional distribution of Θ_i given Y is normal with mean $\mathbf{E}[\Theta_i \mid Y]$ and a certain (conditional) variance which is the same for all possible values of Y. This is a general feature of normal models involving additive normal noise in the observations. In this class, we will not prove this fact in its full generality, but we will see it in some special cases.

[2] An additional argument is needed here to verify that this solution corresponds to a global minimum, as opposed to, say, a saddle point or a local minimum.