Derivation of Conditional Density Equation

Daniel Yamins

June 16, 2003

Derivation of the Conditional Density Equation (CDE) for systems with possibly correlated noise. This is inspired by #9 from p. 111 of the notes for ES 203. We have the system

$$dx = \alpha x dt + dw + budt;\ dy = xdt + d\nu.$$

where $dw, d\nu$ are indepedent Brownian motions and $\alpha \in \{-1, -2\}$. We're suppose to do system identification for α in various circumstances. First, let's assume that $u = 0$: we are now to calculate the conditional probability for propating estimate of α. To do that, we use the unnormalized conditional density equation for the system (α, x). Let $\rho(\alpha, x, t)$ denote the conditional density distribution conditioned on y. Then let us assume that $\rho(-1, x, t)$ and $\rho(-2, x, t)$ are both gaussians for all time, a fair assumption in light of the fact that the initial condition is known to be gaussians for each. Hence, write

$$\rho(-1, x, t) = e^{a_1(t)x^2 + b_1(t)x + c_1(t)}$$

and

$$\rho(-2, x, t) = e^{a_2(t)x^2 + b_2(t)x + c_2(t)}.$$

In this case, the FP operator is:

$$-\frac{\alpha}{2}\frac{\partial\rho}{\partial x} + \frac{1}{2}\frac{\partial^2\rho}{\partial x^2} - \frac{1}{2}x^2$$

so the conditional density equation yields the six equations (three for each value of α)

$$\dot{a}_1 = 2a_1^2 - \frac{1}{2};\ \dot{b}_1 = 2a_1 b_1 - 2a_1 + \frac{dy}{dt};\ \dot{c}_1 = \frac{1}{2}b_1^2 + a_1 - b_1$$

and

$$\dot{a}_2 = 2a_2^2 - \frac{1}{2};\ \dot{b}_2 = 2a_2 b_2 - 4a_2 + \frac{dy}{dt};\ \dot{c}_2 = \frac{1}{2}b_2^2 + a_2 - 2b_2$$

where all these are functions of time. These are the propagation rules for ρ. Using the formulae that we say in class relating the a, b to mean and variance, the initial conditions for the a_i, b_i are set at

$$a_1(0) = a_2(0) = -\frac{1}{2\sigma(0)},$$

$$b_1(0) = b_2(0) = \frac{\hat{x}(0)}{\sigma(0)}$$

where $\hat{x}(0), \sigma(0)$ are the mean and variance of the initial gaussians. The initial conditions for the c_i are set at

$$c_1(0) = c_2(0 = -ln(2)$$

because the a priori probabilities are $\frac{1}{2}$ for both values.

Now,

$$\rho(\alpha,t)=\int\rho(\alpha,x,t)dx.$$

So we can compute this for each value of α easily by "completing the squre" so to speak.

$$\begin{aligned}\int e^{a(t)x^2+b(t)x+c(t)}dx &= e^{c(t)}\int e^{a(t)x^2+b(t)x}dx\\ &= e^{c(t)-\frac{b^2(t)}{4a(t)}}\int e^{-(\sqrt{-a(t)}x-\frac{b(t)}{2\sqrt{-a(t)}})^2}dx\\ &= e^{c(t)-\frac{b^2(t)}{4a(t)}}\frac{1}{\sqrt{-a(t)}}\int e^{-y^2}dy\\ &= \sqrt{\frac{2\pi}{-a(t)}}e^{c(t)-\frac{b^2(t)}{4a(t)}}.\end{aligned}\tag{1}$$

Hence we propagate a_i, b_i, c_i for $i=1,2$ accoring to the six equations above starting at those given initial conditions. Then we compare the value of

$$\sqrt{\frac{2\pi}{-a_i(t)}}e^{c_i(t)-\frac{b_i^2(t)}{4a_i(t)}}$$

at any given time, and whichever one is bigger, that is the one the the data indicants is the rate value of the system.

We can easily see that the mean $\hat{x}_i$ for the value α_i is given by $-\frac{b_i}{2a_i}$. Hence in terms of the a, b, the condition mean is given by

$$\hat{x}(t)=\hat{x}_1+\hat{x}_2=-\frac{1}{2}(\frac{b_1(t)}{a_1(t)}+\frac{b_2(t)}{a_2(t)}).$$

The next thing that we're supposed to do is calculate the conditional probability update equations for α in the case that $u(t)=k(t)z(t)$ where z obeys the stochastic equations

$$dz=-\beta(t)zdt+\gamma(t)dy.$$

This system can be rewritten as

$$\begin{bmatrix}\dot{x}\\ \dot{z}\\ \dot{y}\end{bmatrix}=\begin{bmatrix}\alpha & k(t) & 0\\ \gamma(t) & -\beta(t) & 0\\ 1 & 0 & 0\end{bmatrix}\begin{bmatrix}x\\ z\\ y\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}dw\\ d\nu\end{bmatrix}.$$

What we want to do is find the density for the subsystem

$$\begin{bmatrix}\dot{x}\\ \dot{z}\end{bmatrix}=\begin{bmatrix}\alpha & k(t)\\ \gamma(t) & -\beta(t)\end{bmatrix}\begin{bmatrix}x\\ z\end{bmatrix}+\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}dw\\ d\nu\end{bmatrix}$$

conditioned on observation by y.

We cannot use the usual conditional density equation for this system, because there is a correlation between the noise in the signal and the noise in the observation (that is, there is $d\nu$ in both). So we will take a slightly alternate approach.

Let $\rho(\alpha,x,z,y)$ be the density distribution. Then

$$\rho(\alpha,x,z|y)=\frac{\rho(\alpha,x,z,y)}{\rho(y)}.$$

Hence

$$\rho(\alpha)=\int\int\frac{\rho(\alpha,x,z,y)}{\rho(y)}dxdz.$$

One way to solve this problem is to compute the joint distribution $\rho(\alpha, x, y, z)$, then integrate out x, y, sum up over α to compute $\rho(y)$, then perform the operation given above. We can compute the joint distribution by solving according to the Fokker Planck equation for the joint system:

$$\begin{aligned}\rho_t(\alpha, x, z, y) &= -\alpha(xR)_x - bk(t)z\rho_x - \gamma(t)x\rho_z + \beta(t)(zR)_z + x\rho_y + \frac{1}{2}(\nabla^2\rho + 2\rho_{yz}) \\ &= (\beta(t) - \alpha)\rho - (\alpha + bk(t)z)\rho_x + (\beta(t)z - \gamma(t)x)\rho_z + x\rho_y\frac{1}{2}(\nabla^2\rho + 2\rho_{yz}).\end{aligned} \tag{2}$$

This procudure is to propagate the FP equation for the joint system, then integrate, divide, and integrate to find the formula for $\rho(\alpha)$. However, we can also derive a conditional density type equation for the case of correlated noise. I will show how to do this for a scalar system, but the result will be the same for any system. To do this, consider

$$\dot{x} = f(x)dt + g(x)dw_1 + c_1 dw_2$$

observed by

$$dy = h(x)dt + \chi dv + c_2 dw_2.$$

Then

$$\rho(x|y(t)) = \frac{\rho(x, y(t))}{\rho(y)(t)}.$$

We will derive an unnormalized differential equation for $\rho(x|y(t))$. That is, we will not use the normalization by $\rho(y)(t)$. So first off by the chain rule

$$\frac{\partial\rho(x, y(t))}{\partial t} = \frac{\partial R(x, y)}{\partial t} + \frac{\partial R(x, y)}{\partial y}\frac{dy}{dt}$$

in which $\frac{dy}{dt}$ signifies the Stratonovic differential (remember, with stratonovic calculus, we can use regular calculus rules such as the chain rule!). We use Fokker-Plank to expand the first term, changing notation for ease.[1] This yields:

$$\begin{aligned}\frac{\partial\rho(x, y(t))}{\partial t} &= L_{x,y}\rho + \rho_y\frac{dy}{dt} \\ &= L_x[\rho] - h(x)R_y + \frac{1}{2}(\chi^2 + c_2^2)\rho_{yy} + c_1c_2\rho_{xy} + R_y\frac{dy}{dt}.\end{aligned} \tag{3}$$

Now, the crucial point is that

$$\frac{\partial R}{\partial y} = h(x)R$$

as can easily be seen via the same calculation done in the notes for $R(y|x)$ in general. Hence

$$\rho_t = L_x[\rho] - h^2(x)\rho + \frac{1}{2}h^2(x)\rho + ch'(x)\rho + ch(x)\rho_x + h(x)\rho\frac{dy}{dt}$$

which becomes

$$\rho_t = [L_x + h^2(x)(\frac{1}{2}c_2^2 + \frac{1}{2}\chi^2 - 1)]\rho + c_1c_2[h'(x) + h(x)\frac{\partial}{\partial x}]\rho + h(x)\frac{dy}{dt}\rho.$$

Note that this reduces directly to the regular conditional density equation when either c_1 or c_2 is zero, that is, the signal and observation noises are uncorrelated with standard observation noise.

So now we can simply plug our original situation into a similar calculation (now with correlation between

[1] Meaning, using L_f to refer to the FP operator for a system f and notation ρ_x to denote differentiation of $\rho(x, y, \dots)$ with respect to x.

y and z noises) to get

$$\rho_t(\alpha, x, z|y) = [L_{\alpha,x,z} - \frac{1}{2}x^2]\rho + h(x)\rho\frac{dy}{dt} + \rho_{yz} = [L_{\alpha,x,z} - \frac{1}{2}x^2]\rho + x\frac{dy}{dt}\rho + x\rho_z.$$

Now

$$L_{\alpha,x,z} = (\beta(t) - \alpha)\rho - (\alpha + bk(t)z)\rho_x + (\beta(t)z - \gamma(t)x)\rho_z + \frac{1}{2}(\rho_{xx} + \rho_{zz}).$$

So the goal now is to plug into the modified conditional density equation something of the form

$$\rho(\alpha_i, x, z) = exp\{a_i(t)x^2 + b_i(t)xz + c_i(t)z^2 + d_i(t)x + e_i(t)z + f_i(t)\}$$

for $i = 1, 2$. Then as above we get 12 differential equations (6 for each value α_i) relating the $a_i(t), \ldots, f_i(t)$. Then, we integrate out the x, z in the expression and get unnormalized $\rho(\alpha, t)$ equations. Propogating these along the 12 differential equations and comparing, we take which ever one is bigger at any given moment, and that is the value of α indicated by the observations up to that time. Here, I will simply note that this is the procedure, and not actually do the (long and laborious) calculation. However, the result provided here is quite general, and represents a simple and clean way to create the conditional density equation with less restrictive assumptions than those seen in the literature that I am aware of (meaning, Prof. B, his students' papers, the Wonham 1960 paper, and the books on stochastic control and differential equations in the McKay Library).