NOTES ON THE FOKKER-PLANCK EQUATION LECTURE

HENNING SPREKELER

This script is meant to summarize the lecture on the use of the Fokker-Planck equation for the drift-diffusion model for decision making.

1. The drift-diffusion model

The basic elements of the drift-diffusion model for perceptual decision making are a decision variable X, which undergoes a stochastic diffusion process

$$dX = \mu dt + \sigma dW \tag{1}$$

until it reaches one of two decision bounds $X = \pm a$.

The drift-diffusion model is popular for the description of decision processes, because it can explain not only the probability of behavioral choices, but also the shape of the observed reaction time distributions.

One way of calculating statistical properties of the decision making process (such as the probability of a particular choice or the reaction time) is to run many realizations ("trials") of the stochastic differential equation (1) and record the quantities one is interested in. However, this approach is not very elegant and can be rather time-consuming, because it requires the simulation of many trials.

A more efficient way is to use equations not for an individual trial, but rather for a whole *ensemble* of infinitely many trials. To do so, we can characterize the process by a probability density $p(x, t|x_0)$, which characterizes the probability $p(x, t|x_0)\Delta x$ that the decision variable is found in the small (infinitesimal) interval $[x, x + \Delta x]$ at time t, given that it was at x_0 at time $t = 0$. The basic idea is that we can calculate the time evolution of this distribution if we know the time evolution eq. (1) of the decision variable.

2. The Chapman-Kolmogorov equation

So how can we calculate the temporal evolution of the distribution $p(x, t)$? It is convenient to look at this problem in the case of discrete time steps. Then the problem amounts to calculating the distribution $p(x, t+\Delta t|x_0)$ at time $t+\Delta t$ from the distribution at time t. This is not very difficult, because the probability to be at x at time $t + \Delta t$ is simply the probability of being at x' at time t and then transitioning from x' to x in the interval Δt, summed over all possible values of x'. The probability for each of these "paths" from x_0 via x' to x is given by the product of the probability to get from x_0 to x' and the probability of getting from x' to x. Putting this into equations yields the *Chapman-Kolmogorov equation*:

$$p(x, t + \Delta t|x_0) = \int p(x|x'; \Delta t)p(x', t|x_0)dx' \,, \tag{2}$$

where $p(x|x'; \Delta t)$ is the transition probability from x' to x in the small time interval Δt. This equation is valid for Markov processes, i.e., for the case where the probability of getting from x' to x depends only on x and x', and *not* on the value of the decision variable at other moments in time.

Numerically, a simple way of implementing the Chapman-Kolmogorov equation is by a matrix multiplication. To see this, let us discretize the decision variable into a set of bins $[x_i, x_{i+1}[$ that cover the whole range $[-a, a]$, and which all have the same width $\Delta x = x_{i+1} - x_i$. Let us now consider the probability $p_i(t) = p(x_i, t|x_0)\Delta x$ that the decision variable is in the (small) interval $[x_i, x_i + \Delta x]$. The Chapman-Kolmogorov equation (2) then implies that

$$\begin{aligned}
p_i(t + \Delta t) &= p(x_i, t + \Delta t|x_0)\Delta x && (3)\\
&= \int p(x_i|x'; \Delta t)p(x', t|x_0)\Delta x dx' && (4)\\
&\approx \sum_j p(x_i|x_j; \Delta t)p(x_j, t|x_0)\Delta x \Delta x && (5)\\
&\quad \text{(approximating the integral by a sum over intervals)}\\
&= \sum_j \underbrace{p(x_i|x_j; \Delta t)\Delta x}_{T_{ij}}\, p_j(t) && (6)\\
&= \sum_j T_{ij} p_j(t) \,, && (7)
\end{aligned}$$

so the vector of the "new" probabilities $p_i(t + \Delta t)$ is simply the product of a transition matrix T_{ij} with the vector $p_i(t)$ of the "old" probabilities. By means of a successive multiplication with T, we can therefore calculate the distribution for all moments in the future, if we know the distribution at a given moment in time.

3. The Fokker-Planck equation

Mathematically, the Chapman-Kolmogorov equation is difficult to treat, because it is an integral equation. Therefore, it is often convenient to consider the limit where Δt is very small. In this case, the transition probability $p(x|x'; \Delta t)$ very quickly decays as a function of the difference $|x - x'|$, because in small intervals, the decision variable cannot travel a long way.

If the probability distribution $p(x, t|x_0)$ is "smooth" in some sense, we can expand the integrand of the right hand side of the Chapman-Kolmogorov equation in terms of $x - x'$

and get, in the limit $\Delta t \to 0$, the Fokker-Planck equation

$$\frac{\partial}{\partial t} p(x,t|x_0) = -\mu \frac{\partial}{\partial x} p(x,t|x_0) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} p(x,t|x_0)\,, \tag{8}$$

here, μ and σ denote the drift and the noise level in the original stochastic differential equation (1). You will not be asked about the derivation of the Fokker-Planck equation in the exam. If you are interested, anyway, you can find a detailed derivation, and generally more information on the theory of stochastic processes, in the standard books by van Kampen (1992) and Gardiner (1985).

The equation (8) is a simple variant of the *Fokker-Planck equation*, which is valid for the special case of the drift-diffusion model (1). In general, if you have a stochastic differential equation of the form

$$dX = \mu(X,t)dt + \sigma(X,t)dW\,, \tag{9}$$

there is a corresponding Fokker-Planck equation

$$\frac{\partial}{\partial t} p(x,t|x_0) = -\frac{\partial}{\partial x}\left[\mu(x,t)p(x,t|x_0)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x,t)p(x,t|x_0)\right]\,. \tag{10}$$

Note that the difference between these equations and the equations (1, 8) lies in the dependence of the drift μ and the noise strength σ on the decision variable and on time. The effect on the Fokker-Planck equation is that the partial derivatives with respect to x have to be taken after multiplying the distribution $p(x,t|x_0)$ with μ and σ^2, respectively.

4. The flux and reaction time distributions

The Fokker-Planck equation can be written as

$$\frac{\partial}{\partial t} p(x,t|x_0) = -\frac{\partial}{\partial x}\underbrace{\left[\mu(x,t)p(x,t|x_0) - \frac{1}{2}\frac{\partial}{\partial x}\left[\sigma^2(x,t)p(x,t|x_0)\right]\right]}_{=:J(x,t)} = -\frac{\partial}{\partial x} J(x,t)\,, \tag{11}$$

where we defined the *flux* $J(x,t)$, which quantifies the number $J(x,t)\Delta t$ of trajectories $X(t)$ of the decision variable that cross the value $X = x$ (from the left) in the small time interval Δt. The formulation (11) of the Fokker-Planck equation has the form of a *continuity equation*, which states the change in the number $p(x,t|x_0)\Delta x$ of trajectories in a small interval $[x, x+\Delta x]$ must be equal to the difference of how many trajectories are coming in from the left at x (which is $J(x,t)\Delta t$) and how many are leaving the interval to the right at $x + \Delta x$ (which is $J(x+\Delta x,t)\Delta t$). In the limits $\Delta t, \Delta x \to 0$ this leads to the continuity equation

$$\frac{\partial}{\partial t} p(x,t|x_0) = -\frac{\partial}{\partial x} J(x,t)\,, \tag{12}$$

which is true in a very general sense and is a differential form of the statement that no trajectories simply disappear (other than by crossing the decision boundaries).

The flux is particularly interesting if we are interested in the reaction time distribution that is generated by the drift-diffusion model, because decisions are made whenever the boundaries at $\pm a$ are crossed. Therefore, the reaction time distribution is simply the flux across the boundaries $\pm J(\pm a, t)$[1]. If we know the solution of the Fokker-Planck equation, we can therefore readily calculate the reaction time distributions for both the choices.

5. Boundary conditions

The Fokker-Planck equation is a partial differential equation, and therefore requires not only an initial condition, but also boundary conditions in order to yield a unique solution. Because the Fokker-Planck equation is meant to describe the *conditional* probability $p(x,t|x_0)$ of getting from x_0 to x, the initial condition is clearly $p(x,t=0|x_0) = \delta(x - x_0)$.

The boundary conditions for Fokker-Planck equations are a bit more tricky. Clearly, the decision variable cannot get past the boundary, so $p(x,t|x_0)$ has to be zero for $|x| > a$. But what happens at the boundary? The crucial insight is that the decision bounds act as *absorbing boundaries*, because any trajectory that hits the boundary is "taken out of the game", because a decision has been made. Absorbing boundaries lead to the boundary condition that the probability $p(x,t|x_0)$ has to vanish at the boundary. Why is that? Let us consider a trajectory that sits just below the upper bound $x = +a$ at time t. As outlined in the lecture, the noise in the evolution of the decision variable implies that in roughly half of the cases the trajectory moves up and will thus cross the bound in the next time step, and is thus "lost". The other half goes back down and therefore stays within the interval $[-a, a]$. At the same time, there is a chance that other trajectories coming from smaller values than a move in to sit just below the upper bound at time $t + \Delta t$. Can those trajectories compensate the "loss" of trajectories across the boundary and "back down"? If you think about it for a while, you will understand that if the probability distribution was flat at the boundary (i.e. zero derivative), the balance of trajectories leaving to smaller x and those coming back from below would be roughly even. This is not true for all those trajectories that move "up" and cause a decision. Those are lost without compensation, because nobody ever comes back from across the boundary. Putting the pieces together, this means that in every time step, roughly half of the trajectories are "lost". Therefore, the probability at $x = a$ must decrease by a factor of roughly $1/2$ in every time step, leading to an exponential decay. If we send the size of the time steps to zero, this decay will become infinitely fast. The correct boundary condition for the probability distribution at an absorbing boundary is therefore $p(|x| \geq a, t|x_0) = 0$.

Note that the Fokker-Planck equation can also be applied to problems with different boundary conditions. It could for example describe the diffusion of a particle in a box with solid walls. In this case, the above arguments do not hold, because the boundary is not absorbing. Solid walls act as "reflecting boundaries". For those, it is not the probability distribution p that should vanish at the walls (as for absorbing boundaries), but rather the flux J (because no particle can move through the wall).

---

[1] where the $\pm$ accounts for the fact that we are interested in the flux in the positive direction for the boundary $+a$ and in the flux in the negative direction for the boundary $-a$. The sign of the flux determines in which direction the decision variable is moving on average.

References

Gardiner, C. W. (1985). *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer-Verlag, Berlin & New York, 2nd edition.

van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*. North-Holland, Amsterdam, 2nd edition.