Markov processes and their friends

Abstract

We will embark on a journey that mirrors my own when learning the simplest connection between partial differential equations and probability.

This talk is intended to be accessible for different audiences, at the begining of each section there is a small guidance paragraph to lead you where you may want to go.

A few basic things

If you know your basic probability from secondary school you can skip this section and go directly to the next section.

Let us explain our setting with an example of rolling a single fair dice.

We will generally have a probability space denoted by $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the *sample space* this is a set of all possible outcomes in our *experiment*, for the dice example this is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$ which represents the numbers on the faces of the dice. Then, $\mathcal{F}$ is a sigma algebra of the sample space Ω, for this example it can very well be the power set 2^{Ω}, this is called the *event space*. Finally, $\mathbb{P}$ is a *probability measure*, which means that it's a measure on the space Ω such that $\mathbb{P}(\Omega) = 1$, let us say that in this case is just the *relative frequency probability* $\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6}$.

The rolling a dice experiment is quite boring as it is, but it helps to understand that then for any event $A \in \mathcal{F}$, the quantity $\mathbb{P}(A)$ represents the likelihood of happening of event A under our framework. Say for instance that you are interested in the event $A = \{\text{The number showing is even}\} = \{2, 4, 6\}$, you can then compute in this case the probability as

$$\mathbb{P}(A) = \frac{|A|}{6} = \frac{|\{2, 4, 6\}|}{6} = \frac{1}{2}. \tag{1}$$

The concept of conditional probability is paramount to understand Markov processes, so might as well remind you of it now. Say that you have now two events $A, B \in \mathcal{F}$, and you wish to know what will happen if you already have certain information, so for instance you know that B has happened and want to know the likelihood of A given what you already know. Since you already know some information, you are effectively changing changing your probability space, now you know that the only possible outcomes are those of B and so your new probability space must be $(B, \mathcal{G}, \mathbb{Q})$...

...where $\mathbb{Q}(A) = \mathbb{P}(A \mid B) = \frac{P(A \cap B)}{P(B)}$ and $\mathcal{G}$ is a reduced sigma algebra taking into account the occurrence of B, probably even 2^B. When you think about it, a normal probability measure is a particular case of the conditional probability for $B = \Omega$.

Now for the rolling dice thing, imagine that now you have a friend to play your boring dice game, and you roll the dice in such a way that only your friend can see it...idk you roll it into a box or something. Then your friend tells you *"You rolled an even number!"*, you may wonder *"Hmm, what is the probability of the number being a 4"*. You can compute this as

$$P(\{4\} \mid \{2, 4, 6\}) = \frac{|\{4\}|}{|\{2, 4, 6\}|} = \frac{1}{3}. \tag{2}$$

Now that you are an expert in conditional probability let us think about it at a deeper level. We know random variables are nothing more than measurable functions, in particular from the probability space

above $(\Omega, \mathcal{F}, \mathbb{P})$ into some measurable space, normally $\left(\mathbb{R}^d, \mathcal{B}\right)$ where $\mathcal{B}$ is the *Borel sigma algebra* of $\mathbb{R}^d$, we have then a random variable $X : \Omega \to \mathbb{R}$. For the sice example, we may have a RV

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \text{ is even} \\ 1 & \text{if } \omega \text{ is odd} \end{cases}. \tag{3}$$

Random variables have *distributions* which are functions of the form

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B \subset \mathcal{B}\}), \tag{4}$$

and since we care about Borel sets anyway, we have an interface to talk to the distribution which is the *distribution function*

$$F_X(x) = \mathbb{P}_{X((-\infty, x))} = \mathbb{P}(\{\omega \in \Omega : X(\omega) < x\}), \tag{5}$$

obviously such semi-infinite intervals are Borel sets, but in addition it allows us to see the probability of estimates of the random variable concerned.

At this stage we are interested only on the random variables which have a distribution function which is at least piece-wise differentiable, because for such things there exist the so called *density function* $p(x)$ which satisfies

$$F_X(x) = \int_{-\infty}^{x} p(y)dy, \tag{6}$$

for all $x \in \mathbb{R}$.

Just as quick example because we will mention it later, if a random variable X has the density function

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{7}$$

it is called a *Gaussian random variable.*

Now going back to conditional stuff we consider the random variable X with a density function $p(x)$ for all $x \in \mathbb{R}$, then consider the event $A = \{\omega \in \Omega : a \leq X(\omega) \leq b\}$, the *conditional density* is

$$p(x|A) = \begin{cases} 0 & \text{if } x < a \text{ or } b < x \\ \frac{p(x)}{\int_a^b p(y)dy} & \text{if } a \leq x \leq b \end{cases} \tag{8}$$

Finally, for a sequence of random variables $X_1, ..., X_n$ which we make them stay in a common probability space the *joint distribution* is

$$F_{X_1,...,X_n} = \mathbb{P}(\{\omega \in \Omega : X_i(\omega) \leq x_i, i = 1, ..., n\}), \tag{9}$$

and if that sequence of random variables is $X_1, ..., X_n, ...$, and represents the evolution of a probabilistic system over time instants $t_1, ..., t_n$, then this is a stochastic process and we call the totality of its joint distributions

$$\left\{F_{X_{i_1},...,X_{i_j}} : i_j = 1, 2, ... \text{ and } j = 1, 2, ...\right\}, \tag{10}$$

normally denoted by $F_{t_{i_1},...,t_{i_j}}$ its *probability law.* The reason this is a set and not only a function, is that the stochastic process could be sampled at different time instants, so the probability law is accounting for that.

There is a particular type of process we will mention, these are the *Gaussian processes*, for those all of the joint distributions are Gaussian.

A simple example to observe the probability law is the one of independent and identically distributed (i.d.d.) random variables, where any joint distribution is given by

$$F_{t_{i_1},\dots,t_{i_j}}\left(x_{i_1},\dots,x_{i_j}\right) = F\left(x_{i_1}\right)\dots F\left(x_{i_j}\right). \tag{11}$$

Markov processes (chains)

If you know about discrete time Markov chains you can skip to the next section.

We already know what *stochastic processes* are, just a sequence of *random variables* indexed by some variable which we tend to call t because it's normally thought as the time. So, in a practical sense stochastic processes study the evolution of random phenomena with the pass of time.

One of the most common type of stochastic processes are those which don't really care about how one reaches a certain state, but only about the current state and whatever inference we want to make about the future is dependant only on said state of the process.

Let us define those processes first in discrete time, to get some intuition about them.

Def. (Markov chain). Let $\mathcal{S}$ be the *state space* and (λ_i) a probability distribution on $\mathcal{S}$. Let the $p_{ij} := p\left(x_i, x_j\right) = \mathbb{P}\left(X_{k+1} = x_j \mid X_k = x_i\right)$ for any two states $x_i, x_j \in \mathcal{S}$ be the *transition probability from state x_i to x_j in exactly one time step* at the time $k \in \{0, 1, 2, \dots\}$ this has to be such that $p_{ij} > 0$ for all i, j and $\sum_j p_{ij} = 1$ for all i we call the process (X_n) a time homogeneous *Markov chain* if it satisfies the properties

$$\mathbb{P}(X_0 = x_i) = \lambda_i, \tag{12}$$

and

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n) = p_{n,n+1}. \tag{13}$$

The later is often called *Markov property*.

Importantly, notice that the indices $n, n+1$ in $p_{n,n+1}$ refer only to the states x_n, x_{n+1}, not to the time instants, this is why we call it *time homogeneous*. This means that the probability of jumping from state x_i to x_j is the same regardless of the time instant at which we are.

The Markov property is also called *memoryless property* and it tells us that for the process itself every time there is a new step it resets itself and has no idea what just happened…sad for it, but good for us because we can then apply the same rules at any moment.

We will explore an example now, let $\mathcal{S} = \{A, B\}$, then let us have a distribution $\lambda = \{\lambda_A, \lambda_B\}$, indicating that at time $t = 0$ we start at A with probability $\frac{1}{3}$ and at B with probability $\frac{2}{3}$, let also $p_{AA} = 1 - \alpha$, $p_{AB} = \alpha$, $p_{BA} = \beta$ and $p_{BB} = 1 - \beta$ which indicates that being on A we jump onto B with probability α and stay on A with probability $1 - \alpha$, similarly if we are at B.

We tend to collect the transition probabilities in the so called *transition matrix*, for this case we have

$$P = \begin{pmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{pmatrix} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix} \tag{14}$$

It is easy to see that for any time n the process can either be $X_n = A$ or $X_n = B$, and since being on either of those states there are only two things we can do, jumping onto the other state, or staying in the current state. Intuitively the Markov property holds.

It would be quite boring if we only could analyze one step of the process, so the next question is clearly: What if we start at time k and want to give l steps. One question for the example at hand is what is the probability that if we start at A at k we'll be at A at time $k + l$?

Well, first of all, forget about the initial time k. Our process doesn't even know where it is, let us start at 0 and think about where will we be at time l, also to make things more illustrative we'll pick $l = 2$, so what's the probability that starting on A, after two steps we are still on A?

On each step we have two options. Let's say that after the first jump we remain on A, that happens with probability $p_{AA} = \alpha$, the second jump is independent of the first one and so the probability of remaining after two steps is $p_{AA}^2 = (1-\alpha)^2$, the chain of jumps is

$$A \xrightarrow{p_{AA}} A \xrightarrow{p_{AA}} A \tag{15}$$

One can see that the only other possible chain of jumps is from A to B and B to A so, the probability of making it to B after one jump is $p_{AB} = \alpha$, and then the probability that being on B we jump into A is p_{BA}, therefore the probability of jumping first onto B and then to A is $p_{AB}p_{BA} = \alpha\beta$.

$$A \xrightarrow{p_{AB}} B \xrightarrow{p_{BA}} A \tag{16}$$

In total the probability of being on A after two steps is $(1-\alpha)^2 + \alpha\beta$. Notice that this is exactly the first entry of the matrix

$$P^2 = \begin{pmatrix} (1-\alpha)^2 + \alpha\beta & \alpha(1-\beta) + (1-\alpha)\alpha \\ (1-\beta)\beta + \beta(1-\alpha) & (1-\beta)^2 + \beta\alpha \end{pmatrix} \tag{17}$$

In general, the *n-step transition* probabilities are calculated using the n power of the transition matrix.

Thm. *Let (X_n) be a time homogeneous Markov chain with state space $\mathcal{S}$ and transition matrix $P = \left(p_{ij}\right)_{i,j\in\mathcal{S}}$ the n-step transition probability from i to j is the ij-th element of the matrix $P(n) := P^n$.*

Again is worth to notice that if the chain was not time homogeneous, then this result is not possible. In fact, we would instead have matrices $P_1 = P, P_2, ..., P_n$ for different time steps, and then the n-step transition matrix would be $P(n) = P_1P_2...P_n$.

From here, the first interesting result comes, which from the result above is trivial.

Thm. *(Chapman-Kolmogorov equation 1) Let (X_n) be a Markov chain with state space $\mathcal{S}$ and transition matrix $P = \left(p_{ij}\right)_{i,j\in\mathcal{S}}$. For $n, m \in \mathbb{N}$*

$$p_{ij}(n+m) = \sum_{k\in\mathcal{S}} p_{ik}(n)p_{kj}(m). \tag{18}$$

In matrix notation this is $P^{n+m} = P(n+m) = P(n)P(m) = P^nP^m$.

Although surely, this result is obvious when for n-step probabilities being clearly a matrix (possibly infinite dimensional), it's quite useful that we will want to keep it when we go into more complicated cases like continuous processes. And just for reference, analysis and algebra people may recognize this structure as the so called *semi-group property*. Moreover, it's a powerful result which translates as "if

you need to go from x_i to x_j in $n+m$ steps, you can always stop midway at n steps and continue from there".

But first, to illustrate the usefulness of the result above let us introduce one more little concept: before we talked about an initial distribution used to define a Markov process, we denoted it λ which told us the likelihood of the chain starting at cetain states (the distribution of the chain if I may), well nothing stops us, and in fact one would be quite encouraged to explore the distributions of the process at each time t_n, so that we can denote by $p(n) \in \mathcal{S}$ the distribution of the Markov chain at time t_n, so this is defined as

$$p(n) = \left\{p_{x_i}(n)\right\}_{x_i \in \mathcal{S}}, \tag{19}$$

then the distribution or probability vector at time t_{n+1}, $p(n+1)$ satisfies the equation

$$p(n+1) = p(n)P(n) \tag{20}$$

and applying the above recursively and knowing the initial distribution, we can get the distribution at any time as

$$p(n) = \lambda P(1)P(2)...P(n-1) = \lambda P^{n-1}. \tag{21}$$

We can see then that in order to characterize a Markov process we only need of its n-step transition probability $p(n)$ and its initial distribution λ, so what if we can simply encode the initial distribution in the n-step transition probability?

More Markov processes (the continuous time ones)

> This section doesn't have much, just skip it.

We will not talk much about these ones today, but we can quickly say that normally the first step into continuous processes is by considering those which are only continuous on time and have independent increments which are random variables with a common distribution. Examples of those are Poisson processes and jump processes.

For instance the jump processes have a discrete jumping chain which only tells you where are you jumping to but not at what time, so still have a state space which is discrete, but now the times will be continuous and are random variables with an exponential distribution (a continuous distribution) which are normally called holding times.

Poisson processes are also an example of processes with independent icrements, which means that that for any time instants t_i and t_{i+1} for any $i \in \{0, ..., n-1\}$ the random variables $X_{j+1} - X_j$ are all independent. For a Poisson process, said increments have a Poisson distribution.

And yet more Markov processes (diffusion)

> If you already know about diffusion processes, I don't know what to tell you, why are you even reading this?

When we go into the continuous time and continuous space, things become more tricky. One class of processes which provides a useful framework to work on are the *diffusion processes*.

Remember that the only thing we need to characterize a Markov process is it's n-step transition probabilities. So, let us find out something similar here; in general we can take a measurable space

$(S, \mathcal{S})$ which for this talk we will choose to be $(\mathbb{R}, \mathcal{B})$ and a subset $T \subset \mathbb{R}$, a function $(s, x; t, B) \mapsto P(s, x; t, B)$ for $s, t \in T$ such that $s \leq t$, $B \in \mathcal{B}$ and $x \in \mathbb{R}$ is called a Markov transition function if

1. for all fixed s, t, x, the function $B \mapsto P(s, x; t, \cdot)$ is a probability measure on $\mathcal{B}$, and for $s = t$ it's the Dirac measure at the point x, δ_x,
2. for all fixed s, t, B, the function $x \mapsto P(s, \cdot; t, B)$ is measurable on $\mathcal{B}$,
3. and whenever $s, t, u \in T$ such that $s \leq t \leq u$ and for all $x \in \mathbb{R}$ and $B \in \mathcal{B}$ the Chapman-Kolmogorov equation holds, i.e.

$$P(s, x; u, B) = \int_{\mathbb{R}} P(t, y; u, B) P(s, x; t, dy). \tag{22}$$

Now that we have the transition functions defined, a stochastic process $\{X_t\}_{t \in T}$ with values in $\mathbb{R}$ is called a Markov process with the transition function $P(t, x; s, y)$ if for all $t, u \in T$ such that $t \leq u$ and $y \in \mathbb{R}$, the function $P(t, X_t; u, y)$ serves as a conditional probability $P(X_u = y \mid \mathcal{F}_{\leq T})$ with respect to the sigma algebra generated by X_s for $s \leq t$. We need to define it in such a robust way, compared to the discrete case, because now there is no concept of the "next" time instant.

Naturally, the function $P(s, x; t, B)$ can be interpreted as the probability of the process to hit B at time t, given that at time s it was at x. And therefore $P(s, x; t, \cdot)$ are the transition probabilities. We can also, define the so called *transition densities* as $p(s, x; t, y)$ such that

$$P(s, x; t, B) = \int_B p(s, x; t, y) dy, \tag{23}$$

for transition densities, the Chapman-Kolmogorov equation reads

$$p(s, x; t, y) = \int_{\mathbb{R}} p(s, x; \tau, z) p(\tau, z; t, y) dz \tag{24}$$

A Markov process is then called diffusion process if the following limits exist for all $\varepsilon > 0$, $s \geq 0$, $x \in \mathbb{R}$ and well defined functions $\mu(t, x)$ and $\sigma(t, x)$

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y - x| > \varepsilon} p(s, x; t, y) dy = 0, \tag{25}$$

alternatively

$$\lim_{t \downarrow s} \frac{1}{t - s} P(s, x; t, |y - x| > \varepsilon) dy = 0, \tag{26}$$

this prevents the process to have instantaneous jumps, then

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y - x| < \varepsilon} (y - x) p(s, x; t, y) dy = \mu(s, x) \tag{27}$$

or also

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y - x| < \varepsilon} (y - x) P(s, x; t, dy) = \mu(s, x), \tag{28}$$

or in a more probabilistic way

$$\lim_{t \downarrow s} \frac{1}{t - s} \mathbb{E}(X_t - X_s \mid X_s = x) = \mu(s, x), \tag{29}$$

where μ is called the drift of the process which represents the instantaneous rate of change in the mean of the process given $X_s = x$, and finally

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x| < \varepsilon} (y-x)^2 p(s,x;t,y) dy = \sigma^2(s,x), \tag{30}$$

or

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x| < \varepsilon} (y-x)^2 P(s,x;t,dy) = \sigma^2(s,x), \tag{31}$$

or

$$\lim_{t \downarrow s} \frac{1}{t-s} \mathbb{E}\Big((X_t - X_s)^2 \mid X_s = x \Big) = \sigma(s,x) \tag{32}$$

where σ is termed the diffusion coefficient and represents the instantaneous rate of change of the squared fluctuations of the process given that $X(s) = x$.

An example of a diffusion process is the famous *Wiener process*, also known as *Brownian motion*. The Wiener process $W = \{W_t : t \geq 0\}$ is a *Gaussian process* with independent increments for which the following holds:

- It starts at 0 with probability 1, i.e. $\mathbb{P}(\{\omega \in \Omega : W_0 = 0\}) = 1$
- The expectation at any t is zero, i.e. $\mathbb{E}W_t = 0$
- The variance of any increment is the time difference, i.e. $\mathrm{Var}(W_t - W_s) = t - s$

The transition density of the Wiener process is

$$p_W(s,x;t,y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2(t-s)} \right) \tag{33}$$

and then the limit in (27) give $\mu(t,x) = 0$ and the one in (30) gives $\sigma^1(t,x) = 1$.

For the final trick, for any diffusion process one can derive a pair of PDEs which are satisfied by the probability density of the process for appropriate regularity conditions on μ and σ, these are the *Kolmogorov forward equation*

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}[\mu(t,y)p] - \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(t,y)p] = 0 \tag{34}$$

for (s,x) fixed, and the *Kolmogorov backward equation*

$$\frac{\partial p}{\partial s} + \mu(s,x)\frac{\partial p}{\partial x} - \frac{1}{2}\sigma^2(s,x)\frac{\partial^2 p}{\partial x^2} = 0 \tag{35}$$

for (t,y) fixed. If we write the operator $\mathcal{L}_u = \mu(u,x)\frac{\partial}{\partial x} - \frac{1}{2}\sigma^2(u,x)\frac{\partial^2}{\partial x^2}$, then the backwards equation simply reads

$$\frac{\partial p}{\partial t} + L_s p = 0 \tag{36}$$

and $\mathcal{L}_u^* = \frac{\partial}{\partial y}[\mu(u,y)] - \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(u,y)]$ is the *formal adjoint* of L_u, so the forward equation reads

$$\frac{\partial p}{\partial t} + L_t^* p = 0. \tag{37}$$

Knowing that, we can also see that p_W is the solution to

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial^2 p}{\partial y^2} = 0 \tag{38}$$

and

$$\frac{\partial p}{\partial s} - \frac{1}{2}\frac{\partial^2 p}{\partial x^2} = 0, \tag{39}$$

so in a roundabout way we found out the *fundamental solution to the heat equation.*

The actual proofs that the solution to the PDEs above are the transition probabilities are not seminar friendly, but we can provide a brief justification at least for the forward equation, which is often referred to as the *Fokker-Planck equation.*

Thm. *Suppose that equations (25), (27) and (30) hold locally uniformly on x and the functions μ and σ^2 are locally bounded, then the transition probabilities satisfy the Kolmogorov forward equation (34).*

Proof. Let $f \in C_0^\infty$ and $h = t - s$, then using the definition of derivative

$$\frac{d}{dt}\int_{\mathbb{R}} f(y)P(s,x,t,dy) = \lim_{h\to 0} h^{-1}\int_{\mathbb{R}} f(y)P(s,x,t+h,dy) - \int_{\mathbb{R}} f(z)P(s,x,t,dz), \tag{40}$$

take the right hand side and apply the Chapman-Kolmogorov equation

$$\begin{aligned}\lim_{h\to 0} h^{-1}\int_{\mathbb{R}} f(y)P(s,x,t+h,dy) - \int_{\mathbb{R}} f(z)P(s,x,t,dz) = \\ \lim_{h\to 0} h^{-1}\int_{\mathbb{R}}\int_{\mathbb{R}} (f(y)-f(z))P(t,z,t+h,dy)P(s,x,t,dz),\end{aligned} \tag{41}$$

we compute the innermost integral with respect to y by using the conditions (25 - 30) and the truncated Taylor expansion of f up to the quadratic term

$$\lim_{h\to 0} h^{-1}\int_{\mathbb{R}} (f(y)-f(z))P(t,z,t+h,dy)P(s,x,t,dz) = \frac{1}{2}\sigma^2(t,z)\frac{\partial^2 f}{\partial z^2} + \mu(t,z)\frac{\partial f}{\partial z} \tag{42}$$

since the convergence as $h \to 0$ is uniform in z, we have

$$\frac{d}{dt}\int_{\mathbb{R}} f(y)P(s,x,t,dy) = \lim_{h\to 0}\int_{\mathbb{R}}\left(\frac{1}{2}\sigma^2(t,z)\frac{\partial^2 f}{\partial z^2} + \mu(t,z)\frac{\partial f}{\partial z}\right)P(s,x;t,dz). \tag{43}$$

Where you have inside the formal adjoint of the Fokker-Planck PDE, and therefore you have a measure valued solution of it. ∎

What other things are there?

The material presented introduced the connection between Markov processes and linear parabolic PDEs, but then there are ways to generalize this.

For instance Mckean, 1966 [1] introduces a type of Markov processes which are linked to parabolic equations but in the nonlinear case, although that work is quite general nowadays those processes are mostly known as McKean SDEs, or McKean-Vlasov SDEs.

The nonlinearity introduced is by having a diffusion process whose drift and/or diffusion coefficients depend on the probability law itself, so the nonlinear of the Kolmogorov PDEs above is the Fokker-Planck equations which would generally read

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y}(\mu(t,y,p)p) - \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(t,y,p)p) = 0 \tag{44}$$

Further reading

If you are interested in the topic, you can read about discrete and time continuous Markov processes in Strook's book in Markov processes [2] and the classic book from Doob [3], for exclusively diffusion processes there is Strook & Varadhan book in the topic [4], and finally if you want to see things only from the analytic point of view go to the book on Fokker-Planck equations by Bogachev and his ridiculously OP friends [5].

Bibliography

[1] H. P. McKean, "A class of Markov processes associated with nonlinear parabolic equations," *Proceedings of the National Academy of Sciences*, vol. 56, pp. 1907–1911, 1966, doi: 10.1073/pnas.56.6.1907.

[2] D. W. Stroock, *An introduction to Markov processes*. Springer, 2005.

[3] J. L. Doob, *Stochastic Processes*. John Wiley & Sons, Inc, 1953.

[4] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*. Springer, 1997.

[5] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov, *Kolmogorov-Fokker-Planck Equations*. American Mathematical Society, 2015.