Nelson's early work on probability

William G. Faris

November 24, 2016; revised September 7, 2017

1 Introduction

Nelson's early work on probability treated both general stochastic processes and the particular case of Markov processes. He developed background tools relevant to his important later discoveries, of which perhaps the most spectacular was his Markov field approach to Euclidean field theory. The present account is about the early work in papers [10, 8, 7, 9].

Other papers having significant probability content include [6, 11, 12]. The book [3] is about a Markov process that arises in quantum mechanics. The book [5] is probability in the framework of non-standard mathematics. The book [4] is not about probability, but it has relevant material on one-parameter semigroups of linear operators. The collection [2] has an overview of Nelson's work in various fields.

Since the papers are technical, it may help to fill in some probability background and to give a few illustrative examples. Probability starts with a set Ω. Each point ω in Ω is though of as a possible *outcome* of an experiment. There is also a σ-algebra $\mathcal{B}$ of subsets of Ω. Each set E in $\mathcal{B}$ is an *event*; the event E happens if the outcome ω belongs to E. Finally there is a probability measure μ defined on $\mathcal{B}$. A function f on Ω that is measurable with respect to $\mathcal{B}$ is called a *random variable*. In the following it will be convenient to use the notation μ for a probability measure and also for the corresponding expectation of a random variable, so

$$\langle \mu, f \rangle = \int_\Omega f(\omega)\, \mu(d\omega) \tag{1}$$

is this expectation.

A σ-algebra $\mathcal{B}$ is closed under countable unions, countable intersections, and complements. The σ-algebra used in elementary probability is essentially unique. Consider the situation when Ω is an uncountable complete separable metric space and $\mathcal{B}$ is the Borel σ-algebra (generated by the open subsets). Then every two such measurable spaces $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$ are isomorphic [1][Section 13.1. Borel Isomorphism]. Thus whether the underlying space Ω is taken to be $\mathbf{R}$ or $\mathbf{R}^n$ or $\mathbf{R}^\infty$ or a complete separable metric space of continuous functions, the underlying structure $(\Omega, \mathcal{B})$ is essentially the same. A case could be made that this is the way to do probability: use this one measurable space.

Nevertheless, in more general situations there is a greater variety of examples. In particular, there can be a tension between measure theory and topology. Suppose that Ω is a topological space. For simplicity suppose Ω is compact and Hausdorff (but not necessarily equipped with a metric). The Borel σ-algebra $\mathcal{B}$ is the σ-algebra generated by the open sets. The collection of open sets is closed under finite intersections and arbitrary union. A probability measure is said to be *regular* if

$$\mu(E) = \sup\{\mu(K) \mid K \subseteq E, K \text{compact}\}. \tag{2}$$

For a regular probability measure certain uncountable operations are permitted. Suppose that Γ is a collection of open subsets (possibly uncountable) and $E = \bigcup \Gamma$ is the union of the subsets in Γ. Suppose that whenever $\Gamma_0 \subseteq \Gamma$ is finite, then $\bigcup \Gamma_0$ is in Γ. Then it follows from regularity that

$$\mu(E) = \sup\{\mu(U) \mid U \in \Gamma\}. \tag{3}$$

Nelson made use of uncountable operations in his work on stochastic processes described below. Dudley's book [1][Appendix E: Pathologies of Compact Nonmetric Spaces] presents advantages and disadvantages of this approach.

2 The canonical version of a stochastic process

Here is the topological setting for the construction of a stochastic process. There is a compact Hausdorff topological space X (the space in which the stochastic process takes its values) and a set T (the parameter set). The goal is to construct and study the corresponding *canonical version* of the stochastic process. This is the version in which the probability space is the compact Hausdorff space X^T of all functions from T to X, with the σ-algebra consisting of all Borel subsets, and with a regular probability measure μ defined on these subsets. The stochastic process is thus considered as a random function from T to X.

In a typical example the space X might be the one-point compactification of $\mathbf{R}^n$, and $T = [0, \infty)$. In this case the random function is thought of as a particle path, that is, as a function of time with values in X. However there is nothing preventing consideration of more general parameter spaces.

The technical point is that T may be uncountable. Nevertheless, the space X^T with the product topology is a compact Hausdorff topological space. The Borel sets are generated by the open subsets of the topology, and uncountable unions of open sets are open. The result is that the σ-algebra of Borel sets is large enough to include many interesting sets to which probability may be assigned. In other words, the canonical version gives a natural notion of a stochastic process as a random function.

The basic construction is reasonably simple. Let T be an (uncountable) set. Let X be a compact Hausdorff topological space. Start with a consistent family of regular probability measures μ_S defined on the spaces X^S for $S \subseteq T$ with S finite. The consistency condition is that if $S \subseteq S' \subseteq T$, then the projection of $X^{S'}$ onto X^S sends $\mu_{S'}$ to μ_S. The conclusion is the existence of a regular

probability measure μ defined on X^T such that for each finite S the projection of X^T onto X^S sends μ to μ_S.

E. Nelson, Regular probability measures on function spaces [10] The paper is about the construction of regular probability measures on function spaces and about techniques for proving properties of these measures. A reader may begin by consulting [12][Appendix A], which goes through many of the same ideas for the special case of Wiener measure.

The paper mainly focuses on arguments for controlling fluctuations to prove various continuity properties. Here is a typical situation. Let T be an uncountable set. For each finite set $S \subseteq T$, let U_S be an open set representing the event that a certain fluctuation takes place for the process restricted to S. Suppose also that $S \subseteq S'$ implies $U_S \subseteq U_{S'}$; a fluctuation on a finite set implies the same fluctuation on each larger finite set. Let U be the union of the U_S for finite $S \subseteq T$. Then U is open and so is a Borel set: it is the event that the fluctuation takes place on the uncountable set T. Suppose that each U_S has a small probability. By regularity the set U also has small probability.

The paper [12][Appendix A] gives an example of this technique. A point $\omega : T \to S$ in X^T is a function from time to space. Let $[a, b]$ be a time interval, and let $\epsilon > 0$. For every finite set $S \subseteq [a, b]$ of time points, define the set $U_S = \{\omega \mid \exists s, t \in S \ |\omega(s) - \omega(t)| > 2\epsilon\}$. The set $U = \{\omega \mid \exists s, t \in [a, b] \ |\omega(s) - \omega(t)| > 2\epsilon\}$ is the uncountable union of the sets U_S. Nevertheless, an estimate that bounds the probability of U_S implies an estimate that bounds the probability of U.

Here are two results from the paper. Both of them occur in the following context. The space X is a second countable locally compact Hausdorff space. The parameter set T is also a second countable locally compact Hausdorff space, for instance a time interval $[a, b]$. The measure μ on function space is a regular Borel probability measure on X^T. A point t in T is called a fixed point of discontinuity if the μ probability of the set of paths in X^T that are continuous at t is not equal to one.

For the first result, suppose that λ is a regular Borel measure on T. The assumption of the theorem is that the fixed points of discontinuity of the functions in X^T have λ measure zero. The conclusion is that the set of paths in X^T that are continuous except on a set of λ measure zero has μ probability one.

For the second result,the assumption of the theorem is that there are no fixed points of discontinuity. The conclusion is that the set of paths that are continuous except on a set of first category has μ probability one. (A set of first category is also called a meager set: it is a countable union of nowhere dense sets.) ||

3 Semigroups and Markov processes

A Markov process is a special kind of stochastic process. In the construction of a (time-homogeneous) Markov process, the family of measures μ_S is constructed

by giving an initial point x_0 in X and a family of regular probability measures $E \mapsto p^t(x, E)$ for $E \subseteq X$ indexed by $t \geq 0$ and x in X. This represents the probability that the process has made a transition from a given x to a point in E at elapsed time t. If $S = \{t_1, \ldots, t_n\}$ with $0 \leq t_1 < t_2 < \cdots < t_n$, then the probability μ_S is determined by

$$\int f(\omega(t_1), \ldots, \omega(t_n))\, d\mu_S(\omega) = \tag{4}$$
$$\int \cdots \int p^{t_1}(x_0, dx_1) p^{t_2 - t_1}(x_1, dx_2) \cdots p^{t_n - t_{n-1}}(x_{n-1}, dx_n) f(x_1, \ldots, x_n).$$

The Markov property is reflected in the fact that the probability of making a transition from x_i to E in time $t_{i+1} - t_i$ does not depend on the previous history of how the process got to x_i. The process is time homogeneous because such conditional probabilities depend only on the time difference.

Take $S = \{s + t\}$ and $S' = \{s, s + t\}$. The consistency condition then says that

$$\int p^{s+t}(x, dz) f(z) = \int \int p^t(x, dy) p^s(y, dz) f(z). \tag{5}$$

This may be written in more elegant notation using the definition

$$(P^t f)(x) = \int p^t(x, dy) f(y). \tag{6}$$

Thus $P^t f$ represents the conditional expectation given state x of a function f of the state at later time t. The consistency equation implies that

$$P^{s+t} f = P^t P^s f. \tag{7}$$

This is the semigroup property.

There is a technical issue about what functions f to allow in the semigroup equation. If the functions are defined on a compact Hausdorff space X, then one possible choice is to take all continuous functions.

Example: The randomization process It may help to have an example for which everything may be computed. Fix $\lambda > 0$ and a probability measure ν. Consider the transition probabilities $p^t(x, dy) = \exp(-\lambda t)\delta_x(dy) + (1 - \exp(-\lambda t))\nu(dy)$. A particle starts at x. After time t it is still at x with probability $\exp(-\lambda t)$. Otherwise it has jumped to a random point that is independent of the starting point x. This process has no fixed points of discontinuity, yet the particle paths are discontinuous.

The randomization process has the semigroup

$$P^t f = \exp(-\lambda t) f + (1 - \exp(-\lambda t))\langle \nu, f \rangle. \tag{8}$$

It is easy and instructive to verify the semigroup property. ||

The next example will need the notion of a Gaussian measure. A Gaussian measure γ is determined by its mean and covariance. Here are some basic

definitions. There is a finite-dimensional real vector space V with dual space V'. An element u of V' is a real linear function on V. The value of u in V' on y in V is written as $\langle u, y\rangle$ or as $\langle y, u\rangle$. There is a given point x in V that will be the mean of γ. Thus

$$\langle \gamma, u\rangle = \langle x, u\rangle. \tag{9}$$

Here the left hand side is a pairing between a measure and a linear function, while the right hand side is a pairing between a vector and an element of the dual space.

Furthermore, there is a covariance operator $C : V' \to V$ that defines an inner product on V'. For u in V' define the centered version of u to be $\check{u} = u - \langle \gamma, u\rangle$. Then

$$\langle \gamma, \check{u}\check{v}\rangle = \langle u, Cv\rangle. \tag{10}$$

The left hand side is a pairing between a measure and a quadratic function, while the right hand side is a pairing between an element of the dual space and a vector in the space.

The explicit form of a Gaussian measure is

$$\langle \gamma, f\rangle = \frac{1}{Z_C} \int_V f(y) \exp(-\frac{1}{2}\langle (y-x), C^{-1}(y-x)\rangle)\, dy. \tag{11}$$

Here $C^{-1} : V \to V'$ is the inverse covariance operator, which defines an inner product on V. The constant Z_C is chosen so that $\langle \gamma, 1\rangle = 1$.

This form with the density can be awkward to use. It is often more convenient to calculate with the Fourier transform

$$\langle \gamma, \exp(iu)\rangle = \exp(i\langle x, u\rangle) \exp(-\frac{1}{2}\langle u, Cu\rangle). \tag{12}$$

The left hand side is a pairing of a measure with an exponential function; the right hand side involves pairings between elements of V and V'. In order to prove an assertion about a Gaussian measure and a function f, compute in the special case $f = \exp(iu)$, where u is linear. Then argue that this implies the general case.

Example: The Ornstein-Uhlenbeck process The Ornstein-Uhlenbeck process is a Markov process that is useful and non-trivial, but that also allows explicit computation. This is because the most important random variables are Gaussian, and Gaussian random variables are determined by mean and covariance. The Ornstein-Uhlenbeck process is a simple model of competition between deterministic approach to equilibrium and random diffusion. There are two parameters in the definition of the Ornstein-Uhlenbeck process. The first is a linear transformation $\omega : V \to V$ acting on V. It is assumed that all eigenvalues of ω are in the open right half plane. This is to ensure that for each x in V the vector $\exp(-t\omega)x$ approaches zero as $t \to \infty$. This represents a tendency to damp out the effect of the initial condition x for large time.

The second parameter is a linear transformation $Q : V' \to V$, where V' is the dual space of V. The linear transformation Q is assumed to give an inner

product on V'. Equivalently $Q^{-1} : V \to V'$ gives an inner product on V. The linear transformation Q represents a diffusion constant. There is also a natural damping of the diffusion constant. Let $\omega^\dagger : V' \to V'$ be the adjoint operator. Then the damped diffusion constant is $\exp(-t\omega)Q\exp(-t\omega^\dagger)$.

The Ornstein-Uhlenbeck process at time t is given by a Gaussian random variable with mean $\exp(-t\omega)$ and with covariance given by a covariance operator $C_t : V' \to V$. The covariance is given by

$$C_t = \int_0^t \exp(-r\omega)Q\exp(-r\omega^\dagger)\,dr. \tag{13}$$

It is helpful to think of r as integrating into the past, so that the contribution from r near zero takes into account recent fluctuations, while the contribution from r from near t gives the damped contributions from near the beginning.

If γ_t^x is the Gaussian measure on V with mean $\exp(-t\omega)x$ and covariance C_t, then the Ornstein-Uhlenbeck semigroup is given by

$$(P^t f)(x) = \langle \gamma_t^x, f\rangle. \tag{14}$$

As $t \to \infty$ the mean converges to zero and the the covariance increases to a limit

$$C = \int_0^\infty \exp(-r\omega)Q\exp(-r\omega^\dagger)\,dr. \tag{15}$$

Again r near zero gives the effect of current fluctuations, while r near infinity gives the damped effect of fluctuations from the remote past. This gives the limiting behavior

$$(P^\infty f)(x) = \langle \gamma, f\rangle \tag{16}$$

Here γ is Gaussian with mean zero and covariance C.

The relation between the C_t in the definition of the process and the C that gives its eventual equilibrium is

$$C_t = C - \exp(-t\omega)C\exp(-t\omega^\dagger). \tag{17}$$

Furthermore, by differentiation

$$\omega C + C\omega^\dagger = Q \tag{18}$$

The Ornstein-Uhlenbeck process is fundamental in much of Nelson's later work. In his lectures Dynamical Theories of Brownian Motion [3] it models the velocity of a small particle being knocked around by the molecules in its environment. The diffusive part is due to the thermal fluctuations of the environment, while the dissipative part comes from frictional forces. The balance of these two effects determines the fluctuations of the velocity of the particle as a function of time; it wanders away from zero, but typically not too far.

It also plays a role in Nelson's work on field theory. In Euclidean field theory the starting point for most constructions is the free field. The Ornstein-Uhlenbeck process is the special case of the free field when the parameter space is one-dimensional. ||

4 Invariant measures and detailed balance

If P^t is a semigroup of transition probabilities, then the *adjoint semigroup* $P^{t\dagger}$ is defined on measures by

$$\langle P^{t\dagger} m, g\rangle = \langle m, P^t g\rangle. \tag{19}$$

The map $f \mapsto P^t f$ is called the *backward evolution*, since $(P_t f)(x)$ expresses the the expectation at time t in terms of the initial point x at time zero. The map $m \mapsto P^{t\dagger} m$ is called the *forward evolution*, since it tells how a measure at time zero maps to a measure at time t.

A measure m is an *invariant measure* if $P^{t\dagger} m = m$. Once such a measure is specified, then it is tempting to define an inner product $(f, g) = m(fg)$ and define the *Hilbert space adjoint semigroup* P^{t*} that acts on functions instead of measures. The equation is

$$(P^{t*} f, g) = (f, P^t g). \tag{20}$$

When this equation is written out in detail, it says that

$$\int\int p^{t*}(y, dx) f(x) g(y) m(dy) = \int\int f(x) p^t(x, dy) g(y) m(dx). \tag{21}$$

In brief,

$$p^{t*}(y, dx) m(dy) = p^t(x, dy) m(dx). \tag{22}$$

This says that the adjoint process sends y to x in time t precisely when the original process sends x to y in time t. In other words, the direction of time is reversed.

There is a very special case of this that occurs in many important applications. This is the case of an semigroup that is *self-adjoint* with respect to an invariant measure. The equation is

$$(P^t f, g) = (f, P^t g). \tag{23}$$

The brief form is

$$p^t(y, dx) m(dy) = p^t(x, dy) m(dx). \tag{24}$$

This is called *detailed balance*. It says that the mechanism for getting from y to x is the same mechanism for getting from x to y. There is no need to run time backward.

Example: The randomization process and detailed balance For the randomization process the adjoint semigroup is

$$P^{t\dagger} m = \exp(-\lambda t) m + \langle m, 1\rangle \nu. \tag{25}$$

The invariant measures are multiples of ν, and so ν is the invariant probability measure. The detailed balance equation says that

$$\nu((P^t f) g) = \nu(f(P^t g)) = \langle \nu, f\rangle \langle \nu, g\rangle + \exp(-\lambda t)[\nu(fg) - \nu(f)\nu(g)]. \tag{26}$$

As the randomization takes place the initial correlation fades away exponentially fast. ||

Example: The Ornstein-Uhlenbeck process and detailed balance In this case the restoring drift is determined by $\omega : V \to V$, and the diffusion coefficient is given by $Q : V' \to V$. These data determine the time-dependent covariance operator $C_t : V' \to V$. This covariance converges to the equilibrium covariance $C : V' \to V$ as $t \to \infty$. A Fourier transform calculation shows that the condition for detailed balance is $\exp(-t\omega)C = C\exp(-t\omega^\dagger)$. This may be stated in the equivalent form $\omega C = C\omega^\dagger$, which says that $\omega = C\omega^\dagger C^{-1} : V \to V$ is self-adjoint with respect to the inner product on V defined by C^{-1}. In particular ω has real (and strictly positive) eigenvalues.

It is not hard to show that detailed balance is equivalent to $\omega Q = Q\omega^\dagger$. This says that $\omega = Q\omega^\dagger Q^{-1} : V \to V$ is self-adjoint with respect to the inner product on V defined by Q^{-1}. When detailed balance holds, the conditions involving equilibrium covariance C and diffusion constant Q are closely related, in fact

$$C = \frac{1}{2}\omega^{-1}Q. \tag{27}$$

Equilibrium covariance is proportional to diffusion over damping. ||

Example: The Wiener process The Wiener process (sometimes called Brownian motion process) is the limiting case of the Ornstein-Uhlenbeck process when ω is set equal to zero. There is no restoring force, only a diffusion governed by covariance matrix Q. Start at x and wait for time t; the process is then Gaussian with mean x and covariance Qt. Lebesgue meaure is an invariant measure, but it is not a probability measure. The semigroup is self-adjoint, that is, the process satisfies detailed balance. ||

Example: The Wiener process with constant drift Another revealing example is the process of diffusion with constant velocity motion. This example occurs in Nelson's account [3] of Einstein's physical analysis of Brownian motion. The constant velocity $a \neq 0$ is in the direction of the gravitational field, and it represents the terminal velocity of a particle falling under the influence of gravitation in an environment providing a frictional force. Start at x and wait for time t. The process is then Gaussian with mean $x + at$ and covariance Qt. In this example Lebesgue measure is an invariant measure, but it does not satisfy detailed balance. There is another invariant measure that is more natural, one whose density with respect to Lebesgue measure is $\exp(2aQ^{-1}x)$. This one does satisfy detailed balance. With downward displacement (in the direction of a) the density increases. Picture a cloud of diffusing particles. There are more at lower levels to diffuse upward and fewer at higher levels to diffuse downward. This compensates for the systematic downward drift, leading to detailed balance. ||

E. Nelson, The adjoint Markoff process [8] This paper mainly deals

with semigroups in the framework of discrete time, but the issues are very similar to those in continuous time. It begins with the existence of an invariant measure (not necessarily a finite measure). The adjoint process with respect to an invariant measure is constructed. A main result is the uniqueness (up to a multiple) of the invariant measure of a recurrent process. For the proof, consider two invariant measures. They have a Radon-Nikodym derivative that is a function. This function is used with the adjoint process to define a martingale. The martingale convergence theorem is then used to prove that the function is constant.

A corollary is that a recurrent process can only have one adjoint process. Nelson gives an example of a non-recurrent process with more than one adjoint process. Another example is Wiener measure with constant drift. ||

5 Infinitesimal generators

A Markovian semigroup is *strongly continuous* if for each f the function $t \mapsto P^t f$ is continuous. In this case, the semigroup has an *infinitesimal generator* A. This is a linear transformation defined on a dense set of f with the property that

$$\frac{d}{dt} P^t f = A P^t f \tag{28}$$

for all f in the domain of A. This is called the *backward equation.* There is also a *forward equation*

$$\frac{d}{dt} P^{t\dagger} m = A^\dagger P^{t\dagger} m \tag{29}$$

that acts on measures m. The equation for an invariant measure is $A^\dagger m = 0$. The condition for detailed balance is $m((Af)g) = m(f(Ag))$. This is self-adjointness.

Example: The randomization process and its infinitesimal generator The infinitesimal generator of the randomization process is

$$Af = \lambda(-f + \langle \nu, f \rangle). \tag{30}$$

Its adjoint is

$$A^\dagger m = \lambda(-m + \langle m, 1 \rangle \nu). \tag{31}$$

It is clear that $A^\dagger \nu = 0$. ||

Example: The Ornstein-Uhlenbeck process and its infinitesimal generator In the example of the Ornstein-Uhlenbeck process the infinitesimal generator is

$$Af(x) = \frac{1}{2} Q \nabla^2 f(x) - \langle \omega x, \nabla f(x) \rangle. \tag{32}$$

For each x the second differential $\nabla^2 f(x)$ is a symmetric bilinear form on V, while Q is a symmetric bilinear form on V'. They combine naturally to give a

number. Furthermore, for each x the differential $\nabla f(x)$ is in V'. Since ωx is in V, these pair naturally together to give a number.

The adjoint is

$$A^\dagger \rho(x) = \nabla \cdot \left(\frac{1}{2} Q \nabla \rho(x) + \omega x \rho(x) \right). \tag{33}$$

This operator acts on densities ρ with respect to Lebesgue measure. It has the form of a divergence applied to a vector field. The equation $A^\dagger \rho = 0$ has a solution ρ that is a Gaussian density with $\rho(x)$ proportional to $\exp(-\frac{1}{2} x C^{-1} x)$. This works because the covariance C satisfies $Q = \omega C + C\omega^\dagger = Q$, or $C^{-1} Q C^{-1} = C^{-1}\omega + \omega^\dagger C^{-1}$.

In the case when the Ornstein-Uhlenbeck process satisfies detailed balance, the covariance C satisfies $2\omega C = Q$, or $QC^{-1} = 2\omega$. The solution ρ satisfies the first order vector field equation

$$\frac{1}{2} Q \nabla \rho(x) + \omega x \rho(x) = 0. \tag{34}$$

This is another form of the detailed balance equation. It says that at every x the sum of the diffusive flux $-\frac{1}{2} Q \nabla \rho(x)$ with the drift flux $-\omega x \rho(x)$ is zero. ||

The last section of the paper [8] treats diffusion on a differential manifold X that is compact and connected. The diffusion process is governed by a generator that is a second order linear partial differential operator

$$A = \sum_i \sum_j a^{ij} \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + \sum_i b^i \frac{\partial}{\partial x_i}, \tag{35}$$

where a^{ij} and b^j are smooth functions of the x_k coordinates. This is the sum of a diffusion term and a drift term, but the decomposition depends on the coordinate system.

The matrix a^{ij} is the inverse of a Riemannian metric g_{ij}. With a Riemannian metric there is an associated volume element $\sqrt{g}\, dx$. The differential of a scalar function f is ∇f, while the gradient is the vector field $a \nabla f$. The divergence of a vector field $\mathbf{v}$ is $\nabla \cdot \mathbf{v}$. The divergence depends on the volume element. The generator A has a natural decomposition as a sum. One term is the Laplace-Beltrami operator associated with a. The other first order term comes from a vector field $\mathbf{v}$. The generator is

$$Af = \nabla \cdot a \nabla f + \mathbf{v} \nabla f. \tag{36}$$

Explicitly

$$Af = \sum_i \sum_j \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_i} \sqrt{g} a^{ij} \frac{\partial f}{\partial x_j} + \sum_j v^j \frac{\partial f}{\partial x_j}. \tag{37}$$

(The relation between the drift term and the vector field term is given by the relation $b^j = v^j + \sum_i \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g} a^{ij}}{\partial x_i}$.) The adjoint of A with respect to the volume measure is

$$A^\dagger \rho = \nabla \cdot (a \nabla \rho - \mathbf{v} \rho). \tag{38}$$

If $A^\dagger \rho = 0$, then $\rho\sqrt{g}\,dx$ is the invariant measure. This measure is usually hard to compute explicitly.

There is one case when computation is relatively easy. Suppose that the vector field is the gradient of a scalar field χ, that is,

$$\mathbf{v} = a\nabla\chi, \tag{39}$$

Let $\rho = c\exp(\chi)$. Then $a\nabla\rho - \mathbf{v}\rho = (a\nabla\chi - \mathbf{v})\rho = 0$, so in particular $A^\dagger\rho = 0$. Detailed balance holds, in fact

$$\int (Af)g\,\rho\,\sqrt{g}dx = \int f(Ag)\,\rho\,\sqrt{g}dx = -\int \nabla f\,a\,\nabla g\,\rho\,\sqrt{g}dx. \tag{40}$$

Example: The Ornstein-Uhlenbeck process as a diffusion process Consider the Ornstein-Uhlenbeck process when the detailed balance condition $QC^{-1} = 2\omega$ is satisfied. Take $\chi = -\frac{1}{2}xC^{-1}x$. The gradient is $\frac{1}{2}Q\nabla\chi = -\frac{1}{2}\frac{1}{2}QC^{-1}x = -\omega x$. The linear vector field is the gradient of a quadratic function. ||

E. Nelson, An existence theorem for second order parabolic equations [7] In this paper, X is a connected second-countable differentiable manifold. (For instance, X could be an arbitrary connected open subset of Euclidean space.) Furthermore, A is a second order linear partial differential operator as described above. The coefficients must be sufficiently smooth; otherwise there are no restrictions on the growth of the coefficients.

The problem is to construct a semigroup P^t defined on the space of real bounded measurable functions on X such that for suitable f the functions $P^t f$ satisfy the second order parabolic partial differential equation $\partial P_t f/\partial t = AP^t f$.

The idea of the proof is to first work locally. Consider an open subset that is diffeomorphic to an open ball. First construct a semigroup on a space of continuous functions that vanish on the boundary of the subset. Within the subset the semigroup is generated by the partial differential operator. The corresponding diffusion has the appropriate behavior within the subset, but is killed at the boundary. Piece together the diffusions on these subsets to obtain a diffusion process that works on all of X, at least as long as the diffusing particle stays within X. This diffusion defines the desired semigroup. ||

E. Nelson, Representation of a Markovian semigroup and its infinitesimal generator [9] If a semigroup is strongly continuous, then it has an infinitesimal generator that is defined on a dense subspace [4]. This paper treats two topics related to semigroups on spaces of continuous functions: establishing strong continuity and representing a generator.

Representing a Markovian semigroup The first topic is the representation of a semigroup as a strongly continuous semigroup acting on a space $C(X)$. The starting point is a general representation theorem. Suppose $\mathcal{C}$ is a real commutative C^* algebra. Then $\mathcal{C}$ is isomorphic to $C(X)$ for a locally compact

Hausdorff space. This representation theorem may be applied to C^* algebras of functions that are far from being continuous. One example is all bounded measurable functions on a measurable space. Another is all essentially bounded functions on a measure space.

Suppose there is a Markovian semigroup $t \mapsto P^t$, where each P^t is a linear transformation on $\mathcal{C}$ that preserves order. Let $\mathcal{V}$ be the set of f such that $t \mapsto P^t f$ is strongly continuous. Then $\mathcal{V}$ is a closed subspace of $\mathcal{C}$, and the semigroup acts on $\mathcal{V}$. Unfortunately, $\mathcal{V}$ need not be an algebra. Let $\mathcal{A}$ be the set of f in $\mathcal{V}$ such that $\mathcal{V}$ implies f^2 is in $\mathcal{V}$. Then $\mathcal{A}$ is a closed subalgebra of $\mathcal{C}$ that is contained in $\mathcal{V}$.

Consider the algebraic assumption that $\mathcal{A} = \mathcal{V}$. Then the semigroup acts on the algebra $\mathcal{A}$. In particular it may be represented as a strongly continuous semigroup acting on some $C(X)$.

Representing the infinitesimal generator The second topic is representation of the infinitesimal generator of a semigroup as a sum of diffusion and jump parts. The semigroup may be taken to be acting on $C(X)$ for some compact Hausdorff space X. Suppose the semigroup is strongly continuous, and let $\mathcal{D}$ be the domain of the infinitesimal generator A of the semigroup. The space $\mathcal{D}$ is a Banach space with respect to the graph norm associated with A.

Let $\mathcal{M}$ be the multiplier algebra of $\mathcal{D}$. This is the set of f in $C(X)$ such that $g \in \mathcal{D}$ implies $fg \in \mathcal{D}$. Then $\mathcal{M}$ is a subspace of $\mathcal{D}$. Furthermore, $\mathcal{M}$ is a Banach algebra. There are examples where $\mathcal{M}$ only consists of constant functions. However in the nicest examples $\mathcal{M} = \mathcal{D}$. In the following suppose at least that $\mathcal{M}$ separates points of X.

Consider a point x and the ideal $\mathcal{I}_x$ of functions in $\mathcal{M}$ that have a zero at x. Let $\mathcal{I}_x^k$ of functions in $\mathcal{M}$ that have a zero of order k at x. This means that $\mathcal{I}_x^k$ is in the ideal in $\mathcal{I}_x$ generated by products of k or more elements of $\mathcal{I}_x$.

It may be that the $\mathcal{I}_x^k$ are all equal to $\mathcal{I}_x$. In this case the generator has a representation

$$(Af)(x) = \int_{X \setminus \{x\}} f(y) a(x, dy) \tag{41}$$

for f in $\mathcal{I}_x$. Since $f = f(x) + (f - f(x))$, this is equivalent to

$$(Af)(x) = \int_{X \setminus \{x\}} (f(y) - f(x)) a(x, dy) \tag{42}$$

for f in $\mathcal{M}$. This is the case of pure jumps.

There is a more subtle situation when the generator may also have a diffusion part. This is when the ideals $\mathcal{I}_x^k$ are strictly smaller than $\mathcal{I}_x$. Nelson's first result is that

$$(Ag)(x) = \int_{X \setminus \{x\}} g(y) a(x, dy) \tag{43}$$

for g in $\mathcal{I}_x^3$. This describes the jump part, if present. Nelson's second result is that

$$(Ag)(x) = (\Delta g)(x) + \int_{X \setminus \{x\}} g(y) a(x, dy) \tag{44}$$

for g in $\mathcal{I}_x^2$. This equation defines the second order part Δ of the generator on $\mathcal{I}_x^2$.

It is more awkward to define a first order part of the generator. The space $V_x = \mathcal{I}_x/\mathcal{I}_x^2$ may be considered as the cotangent space at x, and each function $f - f(x)$ determines an element of V_x, the differential of $df(x)$ of f at x. Suppose $f_i - f_i(x)$ for $i = 1, \ldots, n$ define differentials $df_i(x)$ that form a basis for V_x. Write $df(x) = \sum_{i=1}^n a_i df_i(x)$. This says that $f - f(x) \equiv \sum_{i=1}^n (f_i - f_i(x))$ modulo $\mathcal{I}_x^2$. Write

$$f = f(x) + \sum_{i=1}^n a_i(f_i - f_i(x)) + g \tag{45}$$

with g in $\mathcal{I}_x^2$. Then the representation is

$$(Af)(x) = \sum_{i=1}^n a_i(Af_i)(x) + (\Delta g)(x) + \int_{X\setminus\{x\}} g(y)a(x, dy). \tag{46}$$

It is amazing that such a representation is valid in this very general setting. ||

References

[1] Richard M. Dudley, Real Analysis and Probabiity, Cambridge University Press, 2002.

[2] William G. Faris (editor), Diffusion, Quantum Theory, and Radically Elementary Mathematics, Mathematical Notes 47, Princeton University Press, Princeton, NJ 2006.

[3] Edward Nelson, Dynamical Theories of Brownian Motion

[4] Edward Nelson, Topics in Dynamics I

[5] Edward Nelson, Radically Elementary Probability Theory

[6] Edward Nelson, A functional calculus using singular Laplace integrals

[7] Edward Nelson, An existence theorem for second order parabolic equations

[8] Edward Nelson, The adjoint Markoff process

[9] Edward Nelson, Representation of a Markovian semigroup and its infinitesimal generator

[10] Edward Nelson, Regular probability measures on function spaces

[11] Edward Nelson, Les écoulements incompressibles d'énergie finie

[12] Edward Nelson, Feynman integrals and the Schrödinger equation