VI. AFTERTHOUGHTS

1. Two topics that were left out.

THE $\mathcal{L}^2$ SPACE OF A GAUSSIAN PROCESS: Looking back over what I have written, it seems to me that these notes contain almost everything a beginning researcher needs to know about the mathematical basis of Gaussian processes. Somehow, although my original intention was only to cover problems of continuity and extrema, there is a lot of other material that has found its way into the notes. There is one gaping omission, however, and this relates to the $\mathcal{L}^2$ space associated with a Gaussian process, and its Hilbert space structure.

The reason that this material does not appear here is that it has nothing to do with continuity, boundedness, and suprema distributions, and these have been our main concern. Nevertheless, a word of advice to the teacher or the reader: If, after going through these notes you have some time and energy left, read the first seven chapters of Major's (1981) monograph on multiple Wiener integrals, and follow it up with the elegant and powerful central limit theorem of Dynkin and Mandelbaum (1983). Then you will be ready to start a serious reading of the Gaussian literature in all its breadth and beauty.

ON MARKOV PROCESSES AND THE ISOMORPHISM THEOREM: Markov processes have appeared in these notes only in a very peripheral fashion. This is unfortunate, because it tends to reinforce the "well known fact" that Gaussian and Markovian processes have little to do with one another, and this WKF is taking somewhat of a beating at the moment. To explain why, I want to describe for you a simple version of a theorem due to Dynkin (1983). Then I will explain to you why this theorem is interesting.

Let $\{X_t, t \geq 0\}$ be an $\Re^d$-valued, symmetric, Markov process with stationary transition density $p_t(x,y) = p_t(y,x)$. (We really need certain technical side conditions, but shall leave them out here.) Let ξ be an exponential random variable with mean one, independent of X, which we treat as a death time for X, and let Δ be the "cemetary" state for X so that the "killed version" of X is given by the process

$$\hat{X}_t = \begin{cases} X_t, & t < \xi, \\ \Delta, & t \geq \xi. \end{cases}$$

The killed process $\hat{X}_t$ is still a Markov process, with transition density

$e^{-t}p_t(x,y)$, $x,y \in R^d$, and Green's function $g(x,y)$ defined by

$$g(x,y) = \int_0^\infty e^{-t} p_t(x,y)\, dt. \tag{6.1}$$

We shall assume throughout that $0 < g(x,x) < \infty$ for all x.

For each $x_0 \in R^d$ we define the probability P_{x_0} on the space of paths of X, augmented by Δ, via the finite-dimensional distributions

$$\begin{aligned}(6.2)\quad & P_{x_0}\left\{\hat{X}_{t_1} \in B_1, \ldots, \hat{X}_{t_k} \in B_k\right\} \\ &= \frac{1}{g(x_0,x_0)} \int_{B_1} \cdots \int_{B_k} e^{-t_k} p_{t_1}(x_0,x_1) p_{t_2-t_1}(x_1,x_2) \ldots p_{t_k-t_{k-1}}(x_{k-1},x_k) \\ &\qquad \times g(x_k,x_0)\, dx_1 \ldots dx_k\end{aligned}$$

for $t_1, \ldots t_k > 0$ and Borel sets B_i. Properly formulated $P_{x_0}\{\cdot\}$ can be interpreted as $P\{\cdot \mid \hat{X}_0 = x_0, \hat{X}_{\xi-} = x_0\}$, so that P_{x_0} describes a process starting and finishing (after an exponential killing time) at x_0. (Note that at time ξ the path takes the value Δ. Therefore, $P_{x_0}\{\hat{X}_{t_1} \in R^d, \ldots, \hat{X}_{t_k} \in R^d\} = P\{\xi > t_k \mid \hat{X}_0 = x_0, \hat{X}_{\xi-} = x_0\}$.)

We shall be interested in the local time process $\{L_x, x \in R^d\}$ of the killed version $\hat{X}_t$ of X_t, which can be formally expressed as

$$L_x = \int_0^\xi \delta_x(X(t))\, dt, \tag{6.3}$$

where δ_x is the Dirac delta function centered at x. This definition can be made rigorous, as usual, by taking L_x as the Radon-Nikodym derivative of the occupation measure

$$\mu(A) = \int_0^\infty I_A(\hat{X}(t))\, dt,$$

where we follow the convention that for all Borel sets $A \in R^d$, $I_A(\hat{X}_t) = 0$ for $t \geq \xi$. (Another approach is to approximate δ_x by some density function and then pass to the limit.)

To set up Dynkin's theorem, the first thing we have to note is that the

Green's function (6.1) is always positive semi-definite, since

$$\begin{aligned}
&\iint \alpha(x)g(x,y)\alpha(y)\,dxdy \\
&= \int\int\int \alpha(x)e^{-t}p_t(x,y)\alpha(y)\,dxdydt \qquad \text{(by (6.1))} \\
&= \iint\iint \alpha(x)e^{-t/2}p_{t/2}(x,z)\,\alpha(y)e^{-t/2}p_{t/2}(z,y)\,dxdydtdz \\
&\qquad\qquad \text{(by Chapman-Kolmogorov)} \\
&= \iint dzdt\Big\{\int \alpha(x)e^{-t/2}p_{t/2}(z,x)\,dx \int \alpha(y)e^{-t/2}p_{t/2}(z,y)\,dy\Big\} \\
&\qquad\qquad \text{(by symmetry)} \\
&= \iint dzdt\big[E\{\alpha(X_{t/2})|X_0 = z\}\big]^2 \\
&\geq 0.
\end{aligned}$$

Thus g can serve as the covariance function of a centered Gaussian process, $\{G(x), x \in R^d\}$, say. Let (Ω, π) denote the probability space of this process. The measure π is, naturally, completely determined by the covariance function $g(x,y)$. Let $L = \{L_x, x \in R^d\}$ denote the local time process defined in (6.3). This process is defined on a probability space (Ω_1, P_{x_0}) where P_{x_0} is given in (6.2). We denote by E_π and $E_{P_{x_0}}$ expectation with respect to the measures π and P_{x_0} respectively. For reasons we shall explain in a moment, the following result is known as "the isomorphism theorem":

6.1 THEOREM (DYNKIN). *Let F be any positive functional on the space of functions from $\Re^d$ to $\Re$ and let $\xi_x = \frac{G^2(x)}{2}$. Then*

$$E_\pi\Big\{F(\xi)\frac{G^2(x_0)}{g(x_0,x_0)}\Big\} = E_\pi E_{P_{x_0}}\{F(\xi + L)\} \tag{6.4}$$

Equivalently, the process $\xi_\cdot(\omega)$ with the measure $\frac{G^2(x_0)}{g(x_0,x_0)}\pi(d\omega)$ and $\xi_\cdot(\omega) + L(\omega_1)$ with the product measure $\pi(d\omega)P_{x_0}(d\omega_1)$ are equal in distribution.

The importance of this Theorem is that it gives us a way of translating probability statements about certain Markov processes (at least those statements that can be phrased in terms of local time) into statements about Gaussian processes, and *vice versa*. Hence the "isomorphism" in the name of the theorem. As an example, note that it is an immediate consequence of the isomorphism theorem that the local time exists as long as $0 < g(x,x) < \infty$, and that it has a continuous version whenever $\{G(x), x \in R^d\}$ has a version with continuous sample paths. This, of course, is a problem that we now know how to solve. Much deeper information about the sample path properties of the local times of Markov processes can also be deduced from known results on Gaussian processes, many of which are extremely difficult to obtain in a purely Markovian framework. For details, see Marcus and Rosen (1990), who are rewriting and extending previous results on Markov local time via Gaussian tools.

In a very clever application of this approach, Raisa Epstein Feldman (1989) has shown how to take all the hard combinatorics out of calculating hitting time probabilities for random walks with exponential holding times by using Gaussian processes. In Feldman (1990) she has shown how to calculate general exit distributions for Lévy processes via Gaussian techniques, and has obtained Wiener-Hopf factorisation results from the same framework.

The relationship between Markov and Gaussian processes goes even deeper than Theorem 6.1. For example, while Theorem 6.1 is restricted to Markov processes for which the local time exists as a point indexed process, there are many interesting cases where the local time exists only as a distribution. This, of course, makes all the hard work we have done treating Gaussian processes on general parameter spaces even more justified.

Furthermore, by studying additive functionals of Markov processes it is possible to determine conditions for certain set-indexed Gaussian process themselves to have a kind of Markov property (Dynkin (1980)), and the entire $\mathcal{L}^2$ space of certain Gaussian processes defined on spaces of measures and certain generalised Gaussian processes defined on Schwartz space can be characterised via additive functionals of vector Markov processes (Adler and Epstein, (1986)). (This paper also attempts to explain "why" the isomorphism theorem works, and has a reasonable amount of expository material on the general background required for its appreciation.)

A proof of Theorem 6.1 (in a much wider setting) is given in Dynkin (1984a). Other related papers, in which you can find applications and extensions, are Dynkin (1983,1984b) and Adler, Marcus and Zinn (1989). All told, this is an exciting new area of activity, since it is one of the few areas in Probability Theory today that seems to be bringing people of diverse interests together, as opposed to what seems to be the natural trend for mathematicians to become more and more specialised and isolated even from their own colleagues. The reason you should be interested, is that the fact that you got this far into these notes indicates that you have a serious interest in Gaussian processes, and therefore may have something to contribute at this rapidly developing interface.

2. Directions for research.

MAJORISING MEASURES: As I mentioned a long time ago, in the Preface, my original motivation in preparing these Notes was to learn about majorising measures, which, at the time I started work, had only recently come to the fore in the literature of Gaussian processes. At the time, there was much excitement, for Talagrand's wielding of this concept had brought a solution to the fifty year old problem of finding necessary and sufficient conditions for sample path continuity.

Looking back over Talagrand's main result, however, it is clear that we have not progressed quite as much as we might like, for, at this point of time, we still do not understand majorising measures well enough to construct them in general. *If* an entropy condition is satisfied, whether it involves the simple entropy function of Dudley or the two-parameter entropy of Chapter 5, we can build a majorising measure. But, in these cases, we already knew how to handle continuity problems without majorising measures.

The challenge that therefore remains is to either more fully understand majorising measures, or to replace them with a concept that is more amenable to investigation. My personal feeling is that a two-parameter entropy approach will turn out to be the most natural in the long run, but I have been wrong more than once in the past (Wife – personal communication).

In view of Theorem 4.2, it also follows that majorising measures must have much in common with the distribution of the position of the maximum of X, when this is well defined. This is clearly an interesting problem in its own right, and very little is known. It is therefore doubly worthy of study.

EXTREMAL THEORY AND ASYMPTOTIC GROWTH PROPERTIES: To describe this problem, consider the following simple example: Let $X_t, t \in \Re$ be a centered, stationary, Gaussian process with continuously differentiable sample paths. Set

$$M(t) = \sup_{s \in [0,t]} X(s), \tag{6.5}$$

and, for $\lambda > 0$ and Borel $A \subset \Re$

$$N_A(\lambda) = \#\{t \in A : X_t = \lambda,\ X_t' > 0\}, \tag{6.6}$$

$$L_A(\lambda) = \nu(\{t \in A : X_t \geq \lambda\}), \tag{6.7}$$

where $\nu(B)$ is the Lebesgue measure of B. Then $N_A(\lambda)$ is the number of upcrossings of the level λ that X has in A, while $L_A(\lambda)$ is the amount of time, while in A, that X spends above the level λ.

There is an old and rich theory associated with the study of the three limits

$$\lim_{t,\lambda \to \infty} P\{M(t) > \lambda\},$$

as t and λ tend to infinity in a coordinated fashion, and

$$\lim_{x,\lambda \to \infty} N_{xA}(\lambda), \qquad \lim_{x,\lambda \to \infty} L_{xA}(\lambda),$$

where, again, x and λ tend to infinity in a coordinated fashion, $xA = \{xt\colon t \in A\}$, and the limit is to be taken in the sense of weak convergence of random measures.

You can find details of what happens in the limit (for the non-Gaussian, non-differentiable as well as Gaussian differentiable case) in Leadbetter, Lindgren and Rootzén (1983), Leadbetter and Rootzén (1988), and Leadbetter (1987), along with their copious bibliographies. In brief, if X satisfies an appropriate mixing condition, (i.e. $EX_tX_0 \to 0$ fast enough as $t \to \infty$), then the appropriate limiting distribution for M is double exponential, and for N and L the limiting processes are, respectively, Poisson and compound Poisson random measures.

All of these results, however, deal only with Gaussian processes on the real line. There is a rather immediate extension to processes on $\Re^d$, and you can find some older-style results in Adler (1981). There has never been an attempt, however, to extend the extremal theory of Gaussian processes to processes on general parameter spaces.

I leave it to you as a (heavily starred) exercise to see what can be made of this, and what applications would be generated by such results.

LOCAL, DIFFERENTIAL, STRUCTURE OF SAMPLE PATHS: There is a rich theory for Gaussian processes on Euclidean spaces that describes the structure of the sample paths in the vicinity of "rare events" such as local extrema at asymptotically high levels.

For example, let $X(t)$ be a continuously differentiable, stationary Gaussian process on $\Re^k$, with covariance function $R(t)$. Suppose that that at some point τ we have $X(\tau) = u$ and

$$\left.\frac{\partial X(t_1, \ldots, t_k)}{\partial t_i}\right|_{t=\tau} = d_i, \qquad i = 1, \ldots, k.$$

Then, with probability approaching one as $u \to \infty$, the random field has the following *deterministic* representation in a neighbourhood of τ:

$$\text{(6.8)}\quad X(t) = u + \sum_{i=1}^{k} d_i(t_i - \tau_i) + \tfrac{1}{2}u \sum_{i,j=1}^{k} (t_i - \tau_i)\lambda_{ij}(t_j - \tau_j) + O\Big(\frac{1}{u}\Big),$$

where

$$\lambda_{ij} = -\left.\frac{\partial^2 R(t)}{\partial t_i\, \partial t_j}\right|_{t=0}.$$

(For details on a precise formulation of this result see, for example, Leadbetter *et. al.* (1983) for processes on $\Re^1$ or Adler (1981) for random fields.)

Here is an interesting problem. We have developed throughout these notes most of the results that would be needed to extend results like the above to general parameter spaces. However, most parameter spaces will not have the nice differential structure that $\Re^k$ has, which is required to even write down (6.8). What is the extension of (6.8) to general parameter spaces, and how do you prove it? (I think the first question is the more difficult one.)

An answer to this question would, I believe, be of particular interest in the study of set indexed (i.e. empirical related) processes.