My work in probability

—

By Olav Kallenberg, 2021

Here I am cherry-picking some results I have proved through the years that I think are especially interesting or important. Most of the quoted results may be found, often in more general form, in my three Springer books:

(K05) *Probabilistic symmetries and invariance principles*, Springer 2005

(K17) *Random measures, theory and applications*, Springer 2017

(K21) *Foundations of modern probabillity*, 3rd ed., Springer 2021

The following account may illustrate why I consider every area of probability theory "my field." My aim in the Foundations book was to summarize the most important results I know in those different areas. I am sorry for the length of this summary, but I can see no way to explain what I did without providing some details. Still I am omitting most technical definitions and all references to the literature. I should also emphasize that I am only highlighting some of the most interesting results. For a fuller account, the reader needs to go to the three books listed above, where more careful explanations are provided.

Random measures.[1] When Peter Jagers[2] came back from his sabbatical in the US in October 1971, he brought a long paper on random measures and point processes, which became the basis for our regular seminar, where we alternated to speak. Peter also organized a little conference on the subject, which attracted lots of people from Sweden and neighboring Denmark. All this activity inspired me to write a long research paper, which became my Ph.D. thesis that I defended in May 1972.

My theses dealt with uniqueness and convergence criteria for random measures and point processes, including matters of *infinite divisibility* and convergence of *null arrays*. I also considered the theories of *Cox processes* and *thinnings*, and criteria for *invariance in distribution* under *measure-preserving maps*. Finally, I had a substantial section on *Palm distributions*, which is today largely obsolete. Much of this material was closely related to what Klaus Matthes[3] and his many collaborators and students were working on in East-Germany, going into the first edition of their monumental point process monograph, hence their interest in my work.

I should explain the role of random measures in this context. *Point processes* may be regarded as special random measures. More importantly, every discussion of point processes leads naturally into the class of more general

[1]K17, Ch's 2–4, 6; K21, Ch's 15, 30

[2]my mentor and friend in Gothenburg

[3]director of the math institute of the academy of sciences in East-Berlin

random measures. For example, a fundamental role is played by the class of *Cox processes*, which are point processes *directed by* more general random measures. Another example is the case of *conditional intensities* and *Papangelou kernels*, discussed below. As a final example, spatial branching processes lead in the diffusion limit to so-called *super-processes*, which are typically processes of diffuse random measures.

Spatial branching processes.[4] Matthes' interests went far beyond this special problem area, and when he visited Gothenburg he brought with him an open problem on *spatial branching processes*, exhibiting a fundamental dichotomy between the stable and unstable cases, where *stability* essentially means convergence to a steady-state distribution while *unstable* means local extinction. The basic problem is then to find necessary and sufficient conditions for stability. This is where I developed some useful criteria based on temporal and spatial *Palm distributions*, leading to processes that got known as *Kallenberg's backward trees*. Using the latter, it become possible to relate the stability of a spatial branching process to the convergence or divergence of an associated *random walk*.

Statistical mechanics.[5] Another class of problems arose from the work of Fredos Papangelou and others in *stochastic geometry*, related to Rollo Davidson's famous conjecture for stationary *line processes* in the plane (see below). By an ingenious construction, Papangelou showed that, for every sufficiently regular point process, there is an associated *conditional intensity* random measure, now known as the *Papangelou kernel*. This object turned out to have important connections to statistical mechanics and attracted a lot of interest. Using very different methods, based on dual disintegrations of *compound Campbell measures*, I managed to construct a much more general object that I called the *Gibbs kernel*[6]. Here some classical relations in *statistical mechanics*, due to R.L. Dobrushin and others, could be restated in terms of the new kernel. This led me to develop a general theory of conditioning in point processes, involving the dual objects of Gibbs kernels and multivariate Palm measures, in particular involving expressions for *inner* and *outer conditioning*. All this work was first published in 1983.

Stochastic geometry.[7] Through the work of Papangelou, I got interested in processes of *lines* and *flats* in a Euclidean space. Before his death, Rollo Davidson[8] had conjectured that, under a second moment condition, every stationary line process in the plane without pairs of parallel lines is a *Cox process*, defined as a mixture of Poisson processes. Papangelou used

[4]K17, Sec. 13.10; K21, Ch. 30

[5]K17, Ch. 8; K21, Ch. 31

[6]named after the famous mathematical physicist J.W. Gibbs

[7]K17, Ch. 11; K21, Ch. 30

[8]a math prodigy in Cambridge, who died tragically in a climbing accident at age 25

his kernel to reduce the problem to showing that every sufficiently regular, stationary random measure on the space of lines in the plane is a.s. *shift invariant*. Though already Davidson and Klaus Krickeberg, later joined by Papangelou, had managed to prove some theorems of the latter type, Davidson's original conjecture remained elusive. I worked hard for a few years to prove extensions of such invariance theorems, and I also studied extensively the associated *particle systems*. Then one day at Christmas time it suddenly occurred to me that I could give a counterexample to *Davidson's conjecture*, and I spent about a week to write a short paper explaining my argument.

I wasn't sure whether this was really publishable, but I decided anyway to send my note to a journal, and I also mailed copies of my paper to some colleagues working in this area. To explain my hesitation, you need to understand how mathematicians work. For every step in a mathematical program we make *conjectures* suggested by previous knowledge, and the technical work is then to determine whether your guesses are true or false. If true, then you have a theorem, and you are ready to take the next step; if false, you have to discard or modify your conjecture. This is just the daily routine of a mathematician. So, here was a conjecture that turned out to be false: not a big deal, you just had to modify the guess, or else give up on this altogether. To my surprise, this conjecture of Davidson was regarded to be of such a momentous importance, and my counterexample caused a sensation in some mathematical circles.

I should add that the original conjecture dealt with line processes stationary under arbitrary *rigid motions*. One of my key steps was to note that it is equivalent to consider stationarity under the group of *translations*, which made the required technical work so much easier to handle.

Exchangeable sequences and processes.[9] My work on random measures led naturally to a study of *exchangeable sequences* and *processes*. Here there are essentially four different cases, depending on whether the time parameter is discrete or continuous and the length of the time interval is finite or infinite. The case of infinite sequences is described by *de Finetti's theorem*, the case of finite sequences is equivalent to *sampling from a finite population*, and the case of exchangeable processes on $\mathbf{R}_+$ had been described by Bühlmann in 1960, leading to mixtures of *Lévy processes*. It remained to consider the most difficult case of exchangeable processes on $[0,1]$. Here the general representation was stated already in my Ph.D. thesis, and a comprehensive theory of the four cases and their mutual relations was developed in a couple of papers a few months later. At this time I also gave some general conditions for the uniform *convergence of series* of independent processes in the Skorohod space, generalizing some celebrated results of Itô and Nisio.

By the mentioned connection with simple random sampling from a finite population, my work also clarified the asymptotic properties of the re-

[9] K05, Ch's 1, 3; K21, Ch. 27

lated *sampling processes*, and I even wrote a short paper on the relationship between sampling with and without replacement from a finite population, published in a statistics journal. By a strange coincidence, the student Jan Hagberg in Stockholm, an old friend of mine, had also studied the asymptotic behavior of such sampling processes, and some of his results, later published in the Russian probability journal, were closely related to mine.

Lévy processes and stochastic calculus.[10] My work in exchangeability led naturally to a special interest in *Lévy processes*[11]. Indeed, the exchangeable processes on $[0, 1]$ are similar to but more general than the Lévy processes, and it became a challenge to extend the basic path and other properties of Lévy processes to this more general class. Already for the Lévy processes themselves there seemed to be some interesting open problems, and in particular I got interested in extending the classical Itô theory of *stochastic integration* with respect to Brownian motion to the case of general Lévy processes. This I did during my visit to Chapel Hill in 1973–74, which was before Meyer's path-breaking paper on stochastic integration with respect to general *semi-martingales*, and I used mostly the classical approach employed by Itô[12]. In particular, I gave conditions for the *existence* of the integral, derived some basic estimates, and studied especially the special case of *stable integrators*. I also developed a theory for not only the quadratic variation, but for *variations* of arbitrary order.

Though this paper later became partly obsolete through the developments of the Strasbourg school, I considered it at the time to be one of my major accomplishments. Thus, the first time I was invited to give a plenary talk at a big conference, this was the paper I chose to present. Similarly, when I returned to Gothenburg and was invited to give a colloquium talk in the math department, I would speak about my stochastic integration paper. This turned out to be a tough call, for two reasons: 1) even the classical *Itô integration* with respect to Brownian motion, by me regarded as one of the most basic areas of probability theory, was virtually unknown even among probabilists in the audience, and 2) my lecture led to criticism that I should talk instead about point processes, which were supposed to be "my field." The latter labeling remained a curse[13] all through my career.

Predictable sampling.[14] As every gambler knows (or ought to know), in a repeated game of chance such as roulette, you can't improve your chances by using a clever gambling strategy. This common gambling experience was

[10]K05, Ch. 3; K21, Ch's 16, 18, 20, 27

[11]processes with stationary independent increments, named after Paul Lévy, one of the greatest probabilists of all time

[12]famous creator of stochastic calculus

[13]I call it the *zebra syndrome*: once you start in a certain area, you are labeled with some stripes that will never go away.

[14]K05, Ch. 4; K17, Sec. 9.3; K21, Ch's 14, 27

first formalized by Doob[15] as the *optional skipping theorem*, showing that the distribution of an i.i.d. sequence $\xi_1, \xi_2, \dots$ remains the same if you sample the variables ξ_k at any *predictable times* $\tau_1 < \tau_2 < \dots$, so that the sequences (ξ_k) and (ξ_{τ_k}) have the same distribution. This quite elementary result is historically interesting, since Doob's first student was the famous mathematician P.R. Halmos, who got as his Ph.D. problem to study the phenomenon of optional skipping.[16]

One of my most surprising discoveries was the fact that, for any finite or infinite exchangeble sequence (ξ_k) on the index set I and any a.s. distinct predictable times $\tau_1, \tau_2, \dots$ taking values in I, the sequences (ξ_k) and (ξ_{τ_k}) have the same distribution. This leads in particular to a very short proof of Lévy's third (and most difficult) *arcsine law* for Brownian motion. Note that my result doesn't require the sequence (ξ_k) to be infinite, nor the τ_k to be increasing.

A related result in *fluctuation theory* had been proved in 1949 by E. Sparre-Anderson, and Feller writes that S-A's result "was a sensation greeted with incredulity, and the original proof was of an extraordinary intricacy and complexity." Feller goes on to give a simplified proof, which is still quite complicated. With some effort, it is possible to prove a continuous-time version of the cited predictable sampling theorem, which may be stated symbolically in the form

$$\lambda \circ V^{-1} = \lambda \quad \Rightarrow \quad X \circ V^{-1} \stackrel{d}{=} X,$$

where X is an exchangeable process on $I = \mathsf{R}_+$ or $[0, 1]$, λ is Lebesgue measure on I, and V is a *predictable mapping* from I to itself.

Strong stationarity.[17] The three major dependence structures of probability theory are those of *stationarity, Markov processes*, and *martingales* (not to mention the elementary case of mere independence). As every probabilist knows, the Markov property extends, under suitable regularity conditions, to the *strong Markov property*, which simply means that it remains true for every *optional time*[18]. Similarly, Doob's *optional sampling theorem* shows that even the martingale property extends to (suitably bounded) optional times. Then what about *stationarity*, defined as invariance in distribution under arbitrary shifts, does it also extend to arbitrary optional times? The surprising answer is no, and in fact the *strong stationarity* of a random sequence $\xi = (\xi_k)$, in the sense of invariance in distribution under optional shifts, is equivalent to the *de Finetti property*, where the ξ_k are conditionally i.i.d. This observation, a rather easy consequence of Ryll-Nardzewski's characterization, is to me one of the most astonishing facts of modern probability.

---

[15]famous creator of modern martingale theory

[16]This episode is vividly described in Halmos' self-biographical *Authomathography.*

[17]K05, Sec. 2.1; K21, Ch.27

[18]also called *stopping time*

Discounted compensator.[19] For any random measure ξ on a product space $\mathsf{R}_+ \times S$, the *Doob-Meyer decomposition* yields an a.s. unique predictable random measure $\hat{\xi}$ on the same space, called the *compensator* of ξ, such that $\int V d\xi = \int V d\hat{\xi}$ for any predictable process V. When $\xi = \delta_{\tau,\chi}$ for a single random point (τ, χ) and the filtration $\mathcal{F}$ is the one induced by ξ, it is well-known that $\hat{\xi}$ can be calculated from the underlying distribution μ by a simple formula, defining the *natural compensator* of ξ. For general $\mathcal{F}$, this is no longer true, but we can still form the ordinary compensator $\hat{\xi}$. Inserting $\hat{\xi}$ instead of the natural compensator in the mentioned formula, we can solve for μ, which now becomes a random measure ζ that I call the *discounted compensator* of ξ. It is easily seen that ζ can be expressed as a *Doléans exponential*[20]. Much harder is to show that the distribution of the pair (ξ, ζ) is a unique mixture of distributions in the natural case.

The point of this construction is that the discounted compensator has the most powerful mapping properties. For the context, recall that a continuous local martingale M can be be time-changed into a Brownian motion, using the quadratic variation $[M]$ as a new time scale. Similarly, a simple point process ξ with a continuous compensator $\hat{\xi}$ can be time-changed into a homogeneous Poisson process, with $\hat{\xi}$ defining the new time scale. (A third basic case is that of suitable *stable stochastic integrals*, but we don't need to discuss that for the moment.) Now it turns out that the discounted compensator can be used to transform any set of random variables into *independent* variables with *specified distributions*. Here the argument makes full use of the *stochastic calculus* for general semi-martingales, where the main step is to prove that a certain fundamental process is a martingale. The details are too technical to review here, so let me just mention that such mappings can be used to give a short proof of the predictable mapping theorem mentioned earlier. It can also be used to transform simple point processes to Poisson, even when $\hat{\xi}$ is not continuous.[21]

Stochastic differential equations.[22] Consider a *stochastic differential equation* of the form

$$dX_t^i = \sigma_j^i(t, X)\, dB_t^j + b^i(t, X)\, dt, \tag{1}$$

where σ and b are given *progressively measurable* functions and B is a Brownian motion of suitable dimension. Here we distinguish between weak, strong, and functional solutions, where a *strong solution* is a process X satisfying (1) for given B and X_0, a *weak solution* is a pair of processes (X, B) satisfying (1) for given $\mu = \mathcal{L}(X_0)$, and a *functional solution* is a process X satisfying $X = F(X_0, B)$ a.s. for a suitable measurable function F. We also say

[19] K17, Sec. 9.4; K21, Ch. 10

[20] named after he outstanding probabilist Cathérine Doléans-Dade

[21] The continuity condition on $\hat{\xi}$ is known as *quasi-left continuity.*

[22] K21, Ch. 32

that *pathwise uniqueness* holds for (1), if any two solutions X and Y with $X_0 = Y_0$ a.s. will agree a.s., and that *uniqueness in law* holds for (1) if any two solutions X and Y with $X_0 \overset{d}{=} Y_0$ have the same distribution.

In a celebrated 1971 paper, Yamada and Watanabe proved that weak existence and pathwise uniqueness imply strong existence and uniqueness in law. Under the same conditions, they further established the existence of a functional solution, but only in the weaker form $X = F_\mu(X_0, B)$, where the function F_μ may depend on the *initial distribution* $\mu = \mathcal{L}(X_0)$. When studying the proof, I noticed that we do have a functional solution of the stronger form $X = F(X_0, B)$, where the function F is independent of μ. It just takes a couple of pages of tight reasoning to get the stronger result.

Exchangeable and contractable arrays.[23] Around 1980, David Aldous and Douglas Hoover independently[24] established some remarkable coding representations of *exchangeable arrays*. In the two-dimensional case, we then consider arrays $X = (X_{ij})$ of random variables indexed by N^2, and say that X is *separately exchangeable* if the joint distribution $\mathcal{L}(X)$ is invariant under arbitrary permutations p and q in the two indices, *jointly exchangeable* if the same invariance holds under mappings by the common permutations $p = q$. By a clever probabilistic argument, Aldous shows that X is separately exchangeable iff

$$X_{ij} = f(\alpha, \xi_i, \eta_j, \zeta_{ij}), \quad i, j \in \mathsf{N},$$

for a measurable function f on $[0,1]^4$ and some i.i.d. $U(0,1)$ random variables $\alpha, \xi_i, \eta_j, \zeta_{ij}$. (A similar representation in the more general jointly exchangeable case is slightly harder to prove.) Hoover uses a totally different approach, based on subtle ideas from *mathematical logic* and *non-standard analysis*, to prove some similar but more complicated representations for jointly exchangeable arrays of arbitrary dimension. Hoover also gives criteria for two functions f and g to be *equivalent*, in the sense of being useful to represent the same array. Such representations have recently found important applications to *random graphs* and *networks*.

Since I needed the higher-dimensional versions for other purposes (see below), I wanted to find probabilistic proofs of Hoover's general results, locked down in an unpublished manuscript that nobody could read.[25] In view of Ryll-Nardzewski's discovery that, for infinite sequences, exchangeability is equivalent to *contractability*, where all sub-sequences have the same distribution, it also became a challenge to establish contractable versions of the Aldous–Hoover results. After years of effort, I eventually managed to prove that the Hoover-type representations remain valid in the jointly contractable

[23] K 05, Ch. 7; K21, Ch. 28

[24] as far as I know

[25] Hoover's paper, written during a stay at the Institute of Advanced Study in Princeton, is absolutely correct to the last syllable, but nobody I knew was willing to spend years trying to understand the underlying mathematical logic.

case. (The contractable case is in fact much harder, since fewer symmetries are available.) As an immediate consequence, a sub-diagonal array is contractable iff it admits an *extension* to a jointly exchangeable array. Since the latter property doesn't involve the Aldous–Hoover representations, it has long been my primary open problem to give a direct proof of the stated result. Nobody seems to have taken up the challenge.

Multivariate rotational symmetries.[26] The cited paper of David Aldous also contains a representation of two-dimensional, separataly rotatable arrays, where *rotatability* means invariance in distribution under orthogonal transformations of finite sub-arrays. Here already the jointly rotatable case is more difficult, and for years I was looking for representations of jointly rotatable arrays of arbitrary dimension. The problem is naturally discussed in an abstract Hilbert-space setting, where we consider *continuous linear random functionals* (CLRF's) X on tensor products of a *separable Hilbert space*, invariant in distribution under arbitrary *unitary transformations*, applied simultaneously in all "coordinates." The general representations may then be expressed as finite sums of tensor products of *multiple Wiener-Itô integrals*. (Readers familiar with *Malliavin calculus* know that the latter integrals also play a fundamental role to represent the *Malliavin derivative* D and its dual D^*, known as the *divergence operator* or *Skorohod integral*.)

Once the rotatable arrays are well understood, their representations can be used to derive representations of exchangeable or contractable *random sheets* of arbitrary dimension. The resulting formulas are simple and easy to understand, when phrased in terms of WI-integrals on tensor products of suitable Hilbert spaces, but their proofs are extremely difficult, and I consider this entire work to be the most difficult thing I have ever done.[27]

Decoupling and tangential processes.[28] The idea of *decoupling* is that, when studying a pair of random objects (ξ, η), we introduce a pair $(\tilde{\xi}, \tilde{\eta})$, where $\tilde{\xi}$ and $\tilde{\eta}$ are independent with the same distributions as ξ and η. In some special cases, the study of $(\tilde{\xi}, \tilde{\eta})$ may give important information about the distribution of the pair (ξ, η). In particular, this method was used systematically by Jurek Szulga and me when we studied multiple stochastic integrals $X^n f$ for suitable Lévy processes X, by reducing the study to the easier case of integrals $(X_1 \cdots X_n)f$, where $X_1, \ldots, X_n$ are independent with the same distribution as X.

A famous case of decoupling is provided by the *Wald identities* where, under suitable conditions, the mean and variance of a random sum or integral

[26] K05, Ch. 8; K21, Ch's 14, 28

[27] Curiously enough, after publishing those results, along with my functional representation for SDE's, I lost my NSF grant, which was never renewed again. This may give some comfort to readers who feel that they are unfairly treated. Life in math is unfair and erratic. Don't give up!

[28] K17, Sec's 9.6–7, 10.6; K21, Ch's 15–16, 20

$\sum_i V_i X_i$ or $\int V dX$ is the same as if the sequences or processes V and X were independent. Years ago I discovered that the Wald identities can be extended to moments of arbitrary order, provided that the underlying conditions on X and V are correspondingly extended. Such decoupling identities may seem technical and boring, until you become aware that they are in fact powerful enough to give simple alternative proofs of the predictable sampling and related theorems mentioned earlier.

To explain the role of tangential processes, recall that a continuous local martingale M or a ql-continuous simple point process N can be reduced to a Brownian motion or a homogeneous Poisson process, through a *random time-change* based on the quadratic variation $[M]$ or compensator $\hat{N}$, respectively. This suggests that M might be a mixture of Brownian motions with varying rates, or that N is a mixture of non-homogeneous Poisson processes, known as *Cox processes.* Though this is not quite true, we are getting close. In fact, for every continuous local martingale M, we can form a mixture of non-homogeneous Brownian motions $\tilde{M}$, such that M and $\tilde{M}$ have the same quadratic variation $[M] = [\tilde{M}]$. Similarly, for every ql-continuous simple point process N, there exists a Cox process $\tilde{N}$ with the same compensator.[29] For general semi-martingales X, the situation is similar but more complicated. Here the local behavior of X is described by the *local characteristics*, and there exists a semi-martingale $\tilde{X}$ with same local characteristics, such that $\tilde{X}$ is a mixture of processes with *non-random* local characteristics.

The point of the construction is that, although the processes X and $\tilde{X}$ are not identically distributed, they have similar asymptotic properties. Furthermore, the component processes of $\tilde{X}$ have deterministic local characteristics, which means that they have independent increments and can be analyzed by elementary classical methods. In particular, a process with invariant local characteristics is a Lévy process, and we are back to very familiar classical territory. Though entire books have been written about decoupling relations for tangential processes, the previously known results are mostly for discrete time, whereas I am considering the continuous-time case, which contains discrete time as a simple special case.

Invariant Palm measures.[30] Throughout random measure theory, the notion of *Palm measures* plays a crucial role, and various aspects of their theory take up more than a hundred pages in my recent random-measure book, not to mention their role in various applications. To appreciate their importance, we need to recognize that they are essentially extensions of *regular conditional distributions* to general random measures, whose theory goes back to the classical work of Kolmogorov and Doob. Much of my work through the years deals with various aspects of this theory and its applications.

Rather than trying to survey the entire area, I will focus on some very

[29] This procedure I call a *Brownification* of M or a *Coxification* of N.
[30] K17, Ch's 5–8, especially Ch. 7; K21, Ch's 3, 31

special, yet fundamental invariance problems. Technically, given a random measure ξ on S, the Palm measures of ξ form a kernel obtained by disintegration of a so-called *Campbell measure* ρ on a product space $S \times T$, where *disintegration* means that we are slicing up ρ into its contributions to different "lines" $S \times \{t\}$. This may be written as $\rho = \nu \otimes \mu$, where ν is a measure on S called the *supporting measure* and μ is the associated *disintegration kernel* from S to T.

Now suppose that G is a group acting measurably on S and T, such that both ρ and ν are G-invariant. Then a basic problem is to choose a *G-invariant version* of the kernel μ, in the sense that $\mu_{rs} = \theta_r \mu_s$ for any $s \in S$ and $r \in G$, where the θ_r are *shift operators* on T (or the associated measure space). The existence of μ can be proved under general regularity conditions, and early special cases were obtained, in the context Palm measures, in pioneering work by Matthes, Mecke, and others. For most applications, we may take G to be a *locally compact, second countable Hausdorff space* (abbreviated as lcscH).

Now assume instead that the measures ρ and ν are *random* and *jointly stationary* under the action of G. For certain purposes, it is then important to find a G-stationary version of the disintegration kernel μ. This problem is much more difficult, since we may no longer assume ρ and ν to be invariant. Using the fact that *Besicovitch's covering theorem* extends to compact subsets of any *Riemannian manifold*, I managed to prove the conjectured result for any *Lie group G*. To extend the result to the general locally compact case, I needed the subtle fact that every lcscH group G contains an open subgroup H, which is a *projective limit of Lie groups*. In this case, the cosets of H are again open, and the *coset space G/H* is discrete and countable. Even when using these facts, the proof is far from trivial, and completing the proof is one of the accomplishments I am the most proud of.

Particle systems.[31] Let me first describe a version of *Dobrushin's classical theorem* from 1956, which inspired much subsequent work in this area. Here let ξ be a stationary point process on R^d with finite *sample intensity* $\bar{\xi}$ (the asymptotic average of ξ over increasing regions, in the sense of the *multi-variate ergodic theorem*). Further consider some probability measures $\nu_1, \nu_2, \ldots$ on R^d, and let each ξ_n be a *ν_n-transform* of ξ, in the sense that the points of ξ are moved independently according to the distribution ν_n. Then ξ_n converges in distribution toward a Cox process ζ directed by $\bar{\xi}\,\lambda^d$, whenever the ν_n are *asymptotically invariant* in a suitable sense. In fact, the distributional convergence comes in two varieties, weak and strong, and convergence holds in the weak or strong sense depending on whether the asymptotic invariance is weak or strict.

To appreciate this result, we note that, if ν_n equals the n-th *convolution power* μ^{*n} of some fixed distribution μ, then the weak asymptotic invariance

[31] K17, Sec. 5.4; K21, Ch. 30

holds iff μ is *non-lattice*, whereas the strict version holds iff at least one power μ^{*n} is *non-singular*. This yields the desired Cox convergence whenever the points of ξ are moved according to independent *random walks* or *Lévy processes*.

The mentioned result can be deduced from two fundamental theorems, which I have obtained in increasing generality, beginning with a basic limit theorem for *random thinnings*, extending some classical results in this area by Rényi and others. For a more general version, say that the probability kernels ν_n on a space S are *dissipative* if $\sup_t \nu_n(t,B) \to 0$ for every bounded set B. Assuming the ξ_n to be ν_n-transforms of some point processes ν_n, we then get for suitable ξ and η

$$\eta_n \nu_n \xrightarrow{vd} \eta \quad \Leftrightarrow \quad \xi_n \xrightarrow{vd} \xi, \tag{2}$$

in which case ξ is a *Cox process directed by* η, where $\xrightarrow{vd}$ denotes *convergence in distribution* with respect to the *vague topology* on S.

The other basic ingredient is a *smoothing*[32] *theorem* for convolutions of the form

$$\xi * \nu_n \xrightarrow{vP} \bar{\xi}\,\lambda^d \quad \Leftarrow \quad \nu_n \text{ weakly asymptotically invariant}, \tag{3}$$

where $\xrightarrow{vP}$ denotes *convergence in probability* with respect to the vague topology on S. Dobrushin's theorem is an immediate consequence of (2) and (3), but the mentioned results are of course much more general and allow far-reaching applications.

Super-processes.[33] Consider a *spatial branching process* in R^d, where the *life lengths* of the particles are independent and exponentially distributed with rate 2, and each particle either dies or splits into two with equal probability $\frac{1}{2}$. Further assume the spatial movements of the particles to be given by *independent Brownian motions*. Now perform a *scaling*, where the particle density and branching rate are both increased by a factor n, whereas the weight of each particle is reduced by a factor n^{-1}. The spatial motion is not affected by the scaling. As $n \to \infty$, we get in the limit a *measure-valued diffusion process* X, known as a *Dawson-Watanabe super-process* (*DW-process* for short). It turns out that for any fixed $t > 0$, the value X_t is a diffuse random measure of *Hausdorff dimension* 2, and we may think of the process as a randomly evolving diffuse cloud. I feel that the DW-process is the most interesting single object studied in probability theory, and I enjoyed tremendously spending years to contribute to its study.

In some path-breaking papers of Dawson & Perkins, Dynkin, and Le Gall, all published in 1991, the evolution of X was shown to be governed by an underlying discrete branching structure, which makes the process amenable to a detailed probabilistic analysis, involving discrete sets of *ancestors* with associated *clusters*. In particular, for any fixed $t > 0$, the ancestors at an

[32] or ergodic

[33] K17, Sec's 13.1–9; K21, Ch. 13

earlier time $s = t - h$ form a Cox process ζ_s^t directed by $h^{-1}\xi_s$, and the ancestral processes with $s < t$ form an inhomogeneous *Yule branching Brownian motion* approaching ξ_t as $s \to t$. The individual ancestors give rise to i.i.d. clusters, and the resulting cluster structure constitutes a powerful tool for analyzing the process.

In this brief account, it would bring us too far to describe my various results in detail, so let me just mention that much of my work consisted in clarifying the *local structure* of the process, in particular involving a fundamental *Lebesgue approximation*, and further a detailed analysis of the *genealogy* (ancestral structure) in terms of what I call a *Palm tree*. In this work, some basic *duality relations* for *multivariate Palm distributions* play an essential role, and I am also making use of Alison Etheridge's probabilistic description of the *moment structure*, as well as an extended version of Jean-François Le Gall's ingenious *Brownian snake*.

Stochastic differential geometry.[34] Stochastic calculus on Riemannian manifolds has been considered ever since some early work of Itô, and several excellent expositions exist. Around 1980, the famous mathematicians Laurent Schwartz[35] and Paul-André Meyer caused a sensation by showing that much of the theory can be developed on manifolds with much weaker properties. Thus, continuous[36] semi-martingales can be defined and studied in an arbitrary *differential manifold* S, and for the martingale property we need only a *connection*. It is not until we come to Brownian motion that we need to introduce a *Riemannian metric*. Note that this reverses the standard procedure in Euclidean spaces, where we start with Brownian motion, then move on to general martingales, and finally reach the semi-martingales by adding a drift term.

I got attracted to this more general theory, because of both its simplicity and its beauty.[37] To appreciate this achievement, it is important to note that all definitions and results must be *intrinsic*, in the sense of being independent of any embedding into a Euclidean space and remaining valid under any change of local coordinates. The definition of a *semi-martingale* is obvious and elementary, and given a connection ∇ on S, a semi-martingale X on S is called a *martingale*, if for any smooth function f on S,

$$f(X) \stackrel{m}{=} \tfrac{1}{2}\, \nabla f[X],$$

where $\nabla f[X]$ is the *quadratic variation* process of X associated with the *bilinear form* ∇f. The relation $\stackrel{m}{=}$ denotes equality in R up to a martingale term.

[34] K21, Ch. 35

[35] the creator of distribution theory

[36] This qualification is henceforth omitted, since only continuous processes will be considered below.

[37] In fact, Meyer's first paper on the subject is entitled *Stochastic geometry without tears*.

The next challenge is to define *intrinsic* versions of the *local characteristics* of a semi-martingale X in S. Since X is continuous, we are simply looking for intrinsic versions of the *drift* and *diffusion rates*. An obvious requirement is that the drift rate should vanish precisely when X is a martingale, and since the latter property makes sense only for manifolds with a connection, the same thing will be true for the intrinsic drift rate. My versions of local characteristics with the desired properties appear in the last chapter of my Foundations book.[38]

[38]I had been hoping to discuss this matter with Michel Émery in Strasbourg, who is a leading expert in the area, but then the pandemic came along, making air travel too risky.

Springer

http://www.springer.com/978-3-030-61870-4

Foundations of Modern Probability
Kallenberg, O.
2021, XII, 946 p. 1 illus. In 2 volumes, not available separately., Hardcover
ISBN: 978-3-030-61870-4