Large scale limits of random geometric structures
Some exercises, definitions and facts
(Malente, 8-10/3/2017, by G. Peccati)

Every random element considered below is defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\mathbb{E}$ indicating expectation with respect to $\mathbb{P}$.

Some notations/jargon from the lectures

(a) For $n \geq 1$, we set $[n] := \{1, ..., n\}$.

(b) Given a function in k variables $(x_1, ..., x_k)$ we write $\widetilde{f}$ to indicate the *symmetrization* of f, that is :
$$\widetilde{f}(x_1, ..., x_k) = \frac{1}{k!} \sum_{\sigma} f(x_{\sigma(1)}, ..., x_{\sigma(k)}),$$
where the sum runs over all permutations σ of $[k]$.

(c) Po(λ), $\lambda > 0$ is the one-dimensional Poisson distribution with parameter λ.

(d) For $n \geq 0$, $T_n(x) := \sum_{k=0}^{n} S(n,k)x^k$ is the nth *Touchard polynomial*, where
$$S(n, k) := \#\{ \text{ partitions of } [n] \text{ with exactly } k \text{ blocks } \}.$$

(e) $(A, \mathscr{A})$ is a mesurable space such that A is a Polish space and $\mathscr{A}$ is its associate Borel σ-field. Note that this implies that $\{x\} \in \mathscr{A}$ for every $x \in A$. We write μ to indicate a non-atomic Borel measure on $(A, \mathscr{A})$ ("Borel" means that $\mu(B) < \infty$ for every bounded measurable B, so that in particular μ is σ-finite).

(f) $\mathscr{A}_0 := \{B \in \mathscr{A} : \mu(B) < \infty\}$.

(g) $\mathbf{N}_l(A)$ is the class of measures on $(A, \mathscr{A})$ that take values in $\mathbb{N} \cup \{+\infty\}$ and are locally finite (i.e., finite on every bounded set). We endow $\mathbf{N}_l(A)$ with the σ-field $\mathscr{N}$ generated by all sets of the form
$$\{\nu \in \mathbf{N}_l(A) : \nu(B) = k\}, \quad k = 0, 1, ..., \; B \in \mathscr{A}.$$

(h) η is a Poisson process on $(A, \mathscr{A})$ with intensity (or "control") μ. We write $\widehat{\eta} := \eta - \mu$.

(i) The *finite intensity Poisson process* corresponds to an intensity of the form $\lambda\pi$, where $\lambda > 0$ and π is a probability measure.

(j) The *homogeneous Poisson process* with intensity (or "parameter") λ corresponds to the case $(A, \mathscr{A}) = (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and $\mu = \lambda \times$ Lebesgue.

(k) We write $\text{supp}\,(\eta) := \{x \in A : \eta(\{x\}) > 0)\}$ (in view of our assumptions, this is consistent with the usual definition of support as a closed set).

(l) For every $k \geq 1$, we use the notation $D_0 = D_0(k) := \{(x_1, ..., x_k) \in A^k : x_i \neq x_j, \forall i \neq j\}$ (purely non-diagonal sets).

(m) We write $\mathcal{S}_0 := \mathbb{R}$ whereas, for $k \geq 1$, $\mathcal{S}_k$ indicates the collection of all measurable mappings

$$f : A^k \to \mathbb{R} : (x_1, ..., x_k) \mapsto f(x_1, ..., x_k),$$

such that f is symmetric, bounded and such that there exists a measurable bounded set $C \subset A^k$ verifying $f(x_1, ..., x_k) = 0$, for every $(x_1, ..., x_k) \notin C$. Also, $\mathcal{S} := \cup_{k=0}^{\infty} \mathcal{S}_k$.

(n) For $k \geq 1$ and $f \in \mathcal{S}_k$, we set

$$\begin{aligned} U_k(f) &:= \int_A \cdots \int_A f(x_1, ..., x_k) \mathbf{1}_{D_0}(x_1, ..., x_k) \eta(dx_1) \cdots \eta(dx_k) \\ &= \int_A \cdots \int_A^{\neq} f(x_1, ..., x_k) \eta(dx_1) \cdots \eta(dx_k), \\ I_k(f) &:= \int_A \cdots \int_A f(x_1, ..., x_k) \mathbf{1}_{D_0}(x_1, ..., x_k) \widehat{\eta}(dx_1) \cdots \widehat{\eta}(dx_k) \\ &= \int_A \cdots \int_A^{\neq} f(x_1, ..., x_k) \widehat{\eta}(dx_1) \cdots \widehat{\eta}(dx_k) \end{aligned}$$

Also, $U_0(c) = I_0(c) = c$.

(o) We define, for $k \geq 0$

$$\begin{aligned} \mathcal{U}_k &:= \{U_k(f) : f \in \mathcal{S}_k\} \\ \mathcal{C}_k &:= \{I_k(f) : f \in \mathcal{S}_k\} \\ \mathcal{U} &:= \operatorname{span}\{\mathcal{U}_k : k \geq 0\} = \operatorname{span}\{\mathcal{C}_k : k \geq 0\}. \end{aligned}$$

(p) Given $F(\eta) \in \mathcal{U}$, we write $D_x^+ F(\eta) = F(\eta + \delta_x) - F(\eta)$ (*add-one cost operator*). Given $x \in \operatorname{supp}(\eta)$, we set $D_x^- F(\eta) = F(\eta) - F(\eta - \delta_x)$ (*remove-one cost operator*).

(q) Given

$$G = \sum_{\ell=0}^{k} I_\ell(g_\ell) \in \mathcal{U},$$

the operator L acts on G as follows

$$LG = -\sum_{\ell=0}^{k} \ell I_\ell(g_\ell),$$

whereas the pseudo-inverse L^{-1} is defined in the obvious way.

(r) For every $\alpha \in [0,1]$, and $F, H \in \mathcal{U}$, we introduce the operator

$$\Gamma_\alpha(F,H) = \alpha \int_A (D_x^+ F\, D_x^+ H)\, \mu(dx) + (1-\alpha) \int_A (D_x^- F\, D_x^- H)\, \eta(dx),$$

and we declare $\Gamma_{\frac{1}{2}}$ to be the *carré-du-champ* operator.

(s) Given $f \in \mathcal{S}_p$ and $g \in \mathcal{S}_q$, and integers $0 \le l \le r \le p \wedge q$, we define the *contraction kernel* in $p+q-l-r$ variables as follows (r variables are identified and l are integrated out) :

$$\begin{aligned} & f \star_r^l g(x_1, ..., x_{p+q-r-l}) \\ &= \int_A \cdots \int_A f(z_1, ..., z_l, x_1, ..., x_{r-l}, x_{r-l+1}, ..., x_{p-l}) \times \\ &\qquad\qquad \times g(z_1, ..., z_l, x_1, ..., x_{r-l}, x_{p-l+1}, ..., x_{p+q-r-l}) \mu(dz_1) \cdots \mu(dz_l). \end{aligned}$$

Note that $\widetilde{f \star_r^l g} \in \mathcal{S}_{p+q-r-l}$.

(t) d_W and d_{TV} stand, respectively, for the 1-*Wasserstein* and *total variation* distances between the laws of two random variables.

One often needs the following result.

Proposition 0.1 (Product formula) *Let $f \in \mathcal{S}_p$ and $g \in \mathcal{S}_q$. Then,*

$$I_p(f) \times I_q(g) = \sum_{r=0}^{\min(p,q)} r! \binom{p}{r} \binom{q}{r} \sum_{l=0}^{r} \binom{r}{l} I_{p+q-r-l}.\left(\widetilde{f \star_r^l g}\right) \tag{0.1}$$

This result implies in particular that $\mathbb{E}[I_p(f) I_q(g)] = 0$, if $p \neq q$, and

$$\mathbb{E}[I_p(f) I_q(g)] = p! \langle f, g \rangle_{L^2(\mu^p)}, \quad \text{if } p = q.$$

Proof: (Sketch) One has that

$$I_p(f) I_q(g) = \int_{D_0(p) \times D_0(q)} f(x_1, ..., x_p) g(y_1, ..., y_q) \widehat{\eta}^{p+q}(dx_1, ...dx_p, dy_1, ..., dy_q).$$

We can represent the set $D_0(p) \times D_0(q)$ as a disjoint union of the type

$$D_0(p) \times D_0(q) = \bigcup_{r=0}^{\min(p,q)} A(r)$$

where the set $A(r)$ is a union of r-diagonals, in the sense that each of its elements $(x_1, ...x_p, y_1, ..., y_q)$ verifies the following property : there are exactly r elements of the vector $(x_1, ..., x_p)$ that are repeated in the vector $(y_1, ..., y_q)$. Symmetry considerations, together with the fact that μ has no atoms yield that

$$\begin{aligned}
&\int_{A(r)} f(x_1, ..., x_p)g(y_1, ..., y_q)d\widehat{\eta}^{p+q}(dx_1, ...dx_p, dy_1, ..., dy_q) \\
&= r!\binom{p}{r}\binom{q}{r}\int_{D_0(p+q-2r)} f(x_1, ..., x_r, a_1, ..., a_{p-r}) \times \\
&\qquad \times g(x_1, ..., x_r, b_1, ..., b_{q-r})\eta^r(dx_1, ..., dx_r)\widehat{\eta}^{p+q-2r}(da_1, ..., db_{q-r}),
\end{aligned}$$

and the result follows by observing that, for every symmetric function $\varphi : A^r \to \mathbb{R}$,

$$\begin{aligned}
&\int_{A^r} \varphi(x_1, ..., x_r)\eta^r(dx_1, ..., dx_r) \\
&= \sum_{l=0}^{r}\binom{r}{l}\int_{A^r} \varphi(x_1, ..., x_r)\mu(dx_1)\cdots\mu(dx_l)\widehat{\eta}(dx_{l+1})\cdots\widehat{\eta}(dx_r).
\end{aligned}$$

■

Nota : I consider exercises with a [$\star$] more interesting for Thursday's afternoon session.

Excercise 1 Given a random variable X with finite moments of every order, we define the nth **cumulant** of X ($n = 1, 2, ...$) to be the quantity

$$\kappa_n(X) := (-i)^n \frac{\partial}{\partial z} \log \mathbb{E}[e^{izX}]\Big|_{z=0},$$

with $i^2 = -1$.

1. Show that $\kappa_1(X) = \mathbb{E}(X)$, and $\kappa_2(X) = \mathbf{Var}(X)$. Prove that, if X and Y are stochastically independent, then $\kappa_n(X+Y) = \kappa_n(X) + \kappa_n(Y)$, and deduce from this property that cumulants of order ≥ 2 are translation-invariant, that is : for every $c \in \mathbb{R}$ and every $m \geq 2$, $\kappa_m(X+c) = \kappa_m(X)$
2. Prove that, if $X \sim \mathrm{Po}(\lambda)$, then $\kappa_m(X) = \lambda$, for every $m \geq 1$.
3. Compute the cumulants of a Gaussian random variable with mean μ and variance σ^2.

Excercise 2 Let $X \sim \mathrm{Po}(\lambda)$ ($\lambda > 0$). Compute an explicit expression for

$$\mathbb{E}[e^{tX}], \quad t \in \mathbb{R},$$

and deduce that the law of X is determined by its moments. Conclude by proving the so-called **Chen-Stein Lemma**, that is : *a random variable Z with values in $\mathbb{N}$ has the Poisson distribution with parameter λ if and only if*

$$\mathbb{E}[Zf(Z)] = \lambda\mathbb{E}[f(Z+1)],$$

for every bounded mapping f.

Excercise 3 Let $\{X_i : i \geq 1\}$ be a sequence of independent random variables such that $X_i \sim \text{Po}(\lambda_i)$, where the parameters $\lambda_i > 0$ are such that $\lambda^\star := \sum_{i=1}^\infty \lambda_i < \infty$. Show that the sum $X^\star := \sum_{i=1}^\infty X_i$ exists in $L^2(\mathbf{P})$, and that $X^\star \sim \text{Po}(\lambda^\star)$.

Excercise 4 Let $X \sim \text{Po}(1)$. Show that, for every integer m, the quantity $\mathbb{E}[(X-1)^m]$ coincides with the numbers of partitions β of $[m]$, such that every block of β has at least size 2 (that is, β has no singletons).

Excercise 5 Build an example of a σ-field $\mathcal{C}$ of $[0,1]$, such that there exists a measure ν on $([0,1],\mathcal{C})$ with values in $\{0,1\}$ and such that $\nu(\{x\}) = 0$ for every $x \in \mathbb{R}$.

Excercise 6 Conclude the proof of the existence of a Poisson process.

Excercise 7 Prove Mecke formula for the finite intensity Poisson process. The extension to the general situation is then a standard affair.

Excercise 8 Prove the multivariate Mecke formula (for deterministic kernels) for the finite intensity Poisson process. The extension to the general situation is again standard.

Excercise 9 [$\star$] Use the multivariate Mecke formula to show that, for every $k \geq 1$ and every $f \in \mathcal{S}_k$, one has that $\mathbb{E}[I_k(f)] = 0$.

Excercise 10 [$\star$] Show that, for every $k \geq 1$ and every $f \in \mathcal{S}_k$,

$$D_x^+ U_k(f) = kU_{k-1}\big(f(x,\bullet)\mathbf{1}_{D_0}(x,\bullet)\big),$$

where $D_0 = D_0(k) := \{(x_1,...,x_k) : x_i \neq x_j, \forall i \neq j\}$.

Excercise 11 [$\star$] Prove that

$$\Gamma_{\frac{1}{2}}(F,G) = \frac{1}{2}\{L(FG) - GLF - FLG\}$$

(this explains why we call $\Gamma_{\frac{1}{2}}$ a *carré-du-champ*).

Excercise 12 [$\star$] Prove the following *Poincaré inequalities*, valid for every $F \in \mathcal{U}$ and every $\alpha \in [0,1]$:

$$\mathbb{E}[\Gamma_\alpha(L^{-1}F, L^{-1}F)] \leq \mathrm{Var}(F) \leq \mathbb{E}[\Gamma_\alpha(F,F)].$$

Excercise 13 [$\star$] Prove the following *improved Poincaré inequality*, valid for every $F \in \mathcal{U}$ and every $\alpha \in [0,1]$:

$$\mathrm{Var}(F) \leq \frac{1}{2}\mathbb{E}[\Gamma_\alpha(F,F)] + \frac{1}{2}\int_A \{\mathbb{E}(D_x^+ F)\}^2 \mu(dx).$$

It is interesting to notice that a Gaussian version of such an estimate has played an important role in our recent work on compressed sensing : L. Goldstein, I. Nourdin and G. Peccati : "Gaussian phase transitions and conic intrinsic volumes : Steining the Steiner formula", Annals of Applied Probability 2017, Vol. 27(1), 1-47.

Excercise 14 [$\star$] Prove that, for every $F \in \mathcal{C}_k$, $k \geq 1$:

$$\mathbb{E}\int_A (D_x^+ F)^4 \, \mu(dx) = \frac{3}{k}\mathbb{E}\left[F^2 \Gamma_{\frac{1}{2}}(F,F)\right] - \mathbb{E}\left[F^4\right].$$

Excercise 15 Let $k \geq 2$, consider $f \in \mathcal{S}_k$ and let $F = I_k(f)$. Prove the following formula :

$$(2k)! \|\widetilde{f \star_0^0 f}\|^2_{L^2(\mu^{2k})} = 2\mathbb{E}[F^2]^2 + (k!)^2 \sum_{r=1}^{k-1} \binom{k}{r}^2 \|f \star_r^r f\|^2_{L^2(\mu^{2k-2r})}.$$

(This result is a little bit technical, but useful in the forthcoming Exercise 16. You can find a proof at p. 97-98 of my book with I. Nourdin "Normal Approximations with Malliavin Calculus", Cambridge 2012)

Excercise 16 [$\star$] Let $F = I_k(f) \in \mathcal{S}_k$, in such a way that $L^{-1}F = -k^{-1}F$, and assume for simplicity that $\mathrm{Var}(F) = 1$. Recall the bound proved in the lectures : for N a standard normal random variable,

$$d_W(F,N) \leq \mathrm{Var}(k^{-1}\Gamma_{\frac{1}{2}}(F,F))^{1/2} + \left(\mathbb{E}\int_A (D_x^+ F)^4 \mu(dx)\right)^{1/2}$$

(i) Use Exercises 11 and 15 in order to write explicitly $\mathrm{Var}(k^{-1}\Gamma_{\frac{1}{2}}(F,F))$ in terms of the norms of the projections $\mathrm{proj}(F^2 \,|\, \mathcal{C}_k)$, $k = 1, ..., 2k-1$, and deduce that

$$\mathrm{Var}(k^{-1}\Gamma_{\frac{1}{2}}(F,F)) \leq C(\mathbb{E}(F^4) - 3),$$

where F is an absolute constant.

(ii) Use Exercises 11, 14 and 15 to prove that

$$\mathbb{E}\int_A (D_x^+ F)^4 \mu(dx) \le C(\mathbb{E}(F^4) - 3),$$

where C is an absolute constant.

(iii) Deduce the following special case of the main result in Ch. Döbler and G. Peccati "The fourth moment theorem on the Poisson space" (Preprint, 2017) : for some absolute constant K

$$d_W(F, N) \le K\Big\{\mathbb{E}(F^4) - \mathbb{E}(N^4)\Big\}^{1/2}.$$

One should notice that "fourth moment results" and associated techniques now account for a quite substantial body of work, spanning several domains of theoretical and applied probability, like functional inequalities, concentration estimates, geometry of random fields, random matrices, compressed sensing and many more – see the dedicated webpage `https://sites.google.com/site/malliavinstein/home` for a constantly updated resource. The actual possibility of having an exact fourth moment theorem on the Poisson space (as the one described above) was an open problem for several years.

Remark. We recall the following bound, that one can find e.g. in the paper : R. Lachièze-Rey and G. Peccati (2013). Fine Gaussian fluctuations on the Poisson space, I : contractions, cumulants and random geometric graphs. The Electronic Journal of Probability, 18(32), 1-35. It is a direct application of the product formula (0.1). If $F = \sum_{i=1}^M I_{q_i}(f_i) \in \mathcal{U}_M$ has variance σ^2 and N is a centred standard normal, then

$$d_W\left(\frac{F}{\sigma}, N\right) \le \frac{C}{\sigma^2}\left\{\max_{(*)} \|f_i \star_r^l f_j\|_{L^2(\mu^{q_i+q_j-r-l})} + \max_{i=1,\dots,M} \|f_i\|^2_{L^4(\mu^{q_i})}\right\},$$

where $\max_{(*)}$ runs over all choices of indices such that $1 \le l \le r \le q_i \le q_j$, with $l \ne q_j$, and C is absolute constant depending on $q_1 + \cdots + q_M$.

Excercise 17 [$\star$] **(Edge counting in the Gilbert graph)** For every $\lambda > 0$ we denote by η_λ the homogeneous Poisson process on $\mathbb{R}^d$ with parameter λ, and write $\eta_1 = \eta$. We fix a "window" W given by a compact set such that ∂W has zero Lebesgue measure (taking the unit cube centered at the origin is perfectly fine). We will let W "grow", as $n \to \infty$, by setting

$$W_n := n^{1/d} W,$$

in such a way that $W_1 = W$; for simplicity we can assume that $\mathrm{Leb}\, W = 1$. Now consider a <u>bounded</u> sequence of positive numbers $\{t_n : n \ge 1\}$, and define the *Gilbert graph*

$$\widehat{G}_n = (\widehat{V}_n, \widehat{E}_n), \quad n \ge 1,$$

as follows : $\widehat{V}_n = \text{supp}\,(\eta) \cap W_n$, and $x \sim y$ if and only if $\|x-y\| \in (0, t_n)$, where $\| \bullet \|$ stands for the Euclidean norm. Note that by construction $\widehat{V}_n$ has no loops. For simplicity, we set $s_n := t_n n^{-1/d}$; in what follows we will distinguish among four regimes, as $n \to \infty$:

(R1) $ns_n^d \to 0$ and $n^2 s_n^d \to \infty$;

(R2) $ns_n^d \to \infty$;

(R3) $ns_n^d \to c \in (0, +\infty)$ (termodynamic regime) ;

(R4) $n^2 s_n^d \to c \in [0, \infty)$.

We are interested in understanding the behaviour, as $n \to \infty$, of the edge counting statistic

$$\widehat{\mathcal{E}}_n := \#\widehat{E}_n, \quad n \geq 1.$$

(i) Show that $\widehat{\mathcal{E}}_n$ has the same distribution as $\mathcal{E}_n = \#E_n$, which is defined as the number of edges in the random geometric graph $G_n = (V_n, E_n)$ defined as $V_n = \text{supp}(\eta_n) \cap W$, and $x \sim y$ if and only if $\|x-y\| \in (0, s_n)$.

(ii) Show that, in all four regimes $\mathbb{E}[\mathcal{E}_n] \approx n^2 s_n^d$;

(iii) Write $\mathcal{E}_n$ as an element of $\mathcal{U}_2$, and write its chaotic representation.

(iv) Show that, under (**R1**) and (**R4**), the projection on the second chaos dominates asymptotically, and $\text{Var}(\mathcal{E}_n) \approx n^2 s_n^d$.

(v) Show that, under (**R2**), the projection on the first chaos dominates asymptotically, and $\text{Var}(\mathcal{E}_n) \approx n^3 (s_n^d)^2$.

(vi) Show that, under (**R3**), both projections contribute asymptotically, and $\text{Var}(\mathcal{E}_n) \approx n$.

(vii) Let N be a centred standard normal random variable, and set

$$\mathbf{E}_n := \frac{\widehat{\mathcal{E}}_n - \mathbb{E}(\widehat{\mathcal{E}}_n)}{\text{Var}(\widehat{\mathcal{E}}_n)^{1/2}}, \quad n = 1, 2, ...$$

Show that, under (**R1**)—(**R3**)

$$d_W(\mathbf{E}_n, N) \to 0,$$

by showing that, under (**R1**)

$$d_W(\mathbf{E}_n, N) \ll \frac{1}{\sqrt{n^2 s_n^d}},$$

and under (**R2**)—(**R3**)

$$d_W(\mathbf{E}_n, N) \ll \frac{1}{\sqrt{n}}.$$

(viii) Show that, under (**R4**), the random variable $\mathcal{E}_n$ admits either a trivial or a Poisson limit.

(ix) If you feel like it, you can repeat the same analysis for generic subgraph counting, by replacing the "edge" with any connected graph with k vertices (triangles, arcs, squares, cliques, ...) ; in all cases, the asymptotic behaviour boils down to the analysis of four well-chosen regimes.

(x) Another interesting question is about *joint distributions*, for instance : are triangles and edges asymptotically independent ?

Remark. The Gilbert graph plays a prominent role in the beautiful monograph by M.D. Penrose "Random Geometric Graphs", Oxford (2003).