1B Markov Chains

Learn the theorems. Do the 2 example sheets. Weber's notes and cross-reference Norris' book.

Definition 1.2 We say that $(X_n)_{n\geq 0}$ is a **Markov chain** with initial distribution λ and transition matrix P if for all $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

1. $P(X_0 = i_0) = \lambda_{i_0}$
2. $P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$

Theorem 1.3 $(X_n)_{n\geq 0}$ is $Markov(\lambda, P)$ iff for all $n \geq 0$ and $i_0, \dots, i_n \in I$,

$$P(X_0 = i_0, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} \tag{1.1}$$

Example Sheet 1 Michaelmas 2020 (Bauerschmidt)

Q1 Let $X = (X_n)_{n\geq 0}$ be a Markov chain. Show that, conditioned on $X_m = i$, $Z = (Z_n)_{n\geq 0}$ given by $Z_n = X_{n+m}$ is a Markov chain with starting state i.

For each $j \in I$,

$$P(Z_0 = j) = P(X_m = j) = \delta_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{else} \end{cases}$$

For each $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

$$\begin{aligned}
&P(Z_{n+1} = i_{n+1} | Z_0 = i_0, \dots, Z_n = i_n) \\
&= P(X_{m+n+1} = i_{n+1} | X_m = i_0, \dots, X_{m+n} = i_n) \\
&= \frac{P(X_m = i_0, \dots, X_{m+n} = i_n, X_{m+n+1} = i_{n+1})}{P(X_m = i_0, \dots, X_{m+n} = i_n)} \\
&= \frac{\sum_{j_0,\dots,j_{m-1}} P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n+1} = i_{n+1})}{\sum_{k_0,\dots,k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\
&= \frac{\sum_{j_0,\dots,j_{m-1}} P(X_{m+n+1} = i_{n+1} | X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n) P(X_0 = j_0, \dots, X_{m-1} = j_{m-1},}{\sum_{k_0,\dots,k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\
&= \frac{\sum_{j_0,\dots,j_{m-1}} P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)}{\sum_{k_0,\dots,k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\
&= P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) \frac{\sum_{j_0,\dots,j_{m-1}} P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)}{\sum_{k_0,\dots,k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\
&= P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) = p_{i_n i_{n+1}} = P(Z_{n+1} = i_{n+1} | Z_n = i_n)
\end{aligned}$$

Q2 Let $X = (X_n)_{n\geq 0}$ be a sequence of independent random variables. Show that X is a Markov chain. Under what condition is this chain homogeneous?

Since $X_0, X_1, \dots, X_n$ are independent, for all $i_0, \dots, i_{n+1} \in I$,

$$\begin{aligned}
P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) &= \frac{P(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1})}{P(X_0 = i_0, \dots, X_n = i_n)} \\
&= \frac{P(X_0 = i_0) \dots P(X_n = i_n) P(X_{n+1} = i_{n+1})}{P(X_0 = i_0) \dots P(X_n = i_n)} \\
&= \frac{P(X_n = i_n) P(X_{n+1} = i_{n+1})}{P(X_n = i_n)} \\
&= \frac{P(X_n = i_n, X_{n+1} = i_{n+1})}{P(X_n = i_n)} \\
&= P(X_{n+1} = i_{n+1} | X_n = i_n) = P(X_{n+1} = i_{n+1})
\end{aligned}$$

so X is a Markov chain. It is homogeneous iff the X_i are identically distributed.

Q3 Let $X = (X_n)_{n \geq 0}$ be a sequence of fair coin tosses (with the two possible outcomes interpreted as 0 and 1) and set $M_n = \max_{k \leq n} X_k$. Show that $(M_n)_{n \geq 0}$ is a Markov chain and find the transition probabilities.

$$P(M_{n+1} = 0 | M_0 = i_0, \dots, M_n = i_n) = \begin{cases} 1/2 & \text{if } i_0 = \dots = i_n = 0 \\ 0 & \text{else} \end{cases}$$

and

$$P(M_{n+1} = 0 | M_n = i_n) = \begin{cases} 1/2 & \text{if } i_n = 0 \\ 0 & \text{else} \end{cases}$$

but $i_0 = \dots = i_n = 0$ iff $i_n = 0$ so the two (conditional) probabilities are the same.

$$P(M_{n+1} = 1 | M_0 = i_0, \dots, M_n = i_n) = \begin{cases} 1/2 & \text{if } i_0 = \dots = i_n = 0 \\ 1 & \text{else} \end{cases}$$

and

$$P(M_{n+1} = 1 | M_n = i_n) = \begin{cases} 1/2 & \text{if } i_n = 0 \\ 1 & \text{else} \end{cases}$$

but $i_0 = \dots = i_n = 0$ iff $i_n = 0$ so the two (conditional) probabilities are the same.

In particular, $p_{01} = p_{00} = 1/2$ and $p_{11} = 1$ and $p_{10} = 0$.

Q4

1B Statistics

Do the 3 example sheets. Weber's notes.

Statistics is a collection of procedures and principles for gaining and processing information in order to make decisions when faced with uncertainty. Two of its principal concerns are **parameter estimation** and **hypothesis testing**.

Data refers to a collection of numbers or other pieces of information to which meaning has been attached. The numbers 1.1, 3.0, 6.5 are not necessarily data. They become so when we are told that they are the muscle weight gains in kg of three athletes who have been trying a new diet.

In statistics, our data are modelled by a vector of random variables

$$X = (X_1, X_2, \ldots, X_n)$$

where X_i takes values in $\mathbb{Z}$ or $\mathbb{R}$.

A **statistic** $T(x)$ is any function of the data. An **estimator** of a parameter θ is a function $T = T(X)$ which we use to estimate θ from an observation of X. T is said to be **unbiased** if

$$\mathbb{E}(T) = \theta$$

The expectation above is taken over X. Once the actual data x is observed, $t = T(x)$ is the **estimate** of θ obtained via the estimator T.

Suppose that the random variable X has probability density function $f(x|\theta)$. Given the observed value x of X, the **likelihood** of θ is defined by

$$lik(\theta) = f(x|\theta)$$

Thus we are considering the density as a function of θ, for a fixed x. In the case of multiple observations, i.e. when $x = (x_1, \ldots, x_n)$ is a vector of observed values of $X_1, \ldots, X_n$, we assume unless otherwise stated that $X_1, \ldots, X_n$ are IID; in this case $f(x_1, \ldots, x_n|\theta)$ is the product of the marginals,

$$lik(\theta) = f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

It makes intuitive sense to estimate θ by whatever value gives the greatest likelihood to the observed data. Thus the **maximum likelihood estimate** $\hat{\theta}(x)$ of θ is defined as the value of θ that maximises the likelihood. Then $\hat{\theta}(X)$ is called the **maximum likelihood estimator (MLE)** of θ.

Of course the maximum likelihood estimator need not exist, but in many examples it does. In practice, we usually find the MLE by maximising $\log f(x|\theta)$, which is known as the **loglikelihood**.

1B Optimisation

Learn the simplex algorithm. Do the 2 example sheets. Weber's notes.

Part 2 Probability & Measure

The relationships between various modes of convergence: almost sure, in probability, in L^p. Strong law of large numbers and central limit theorem.

Part 2 Applied Probability

Review statements and proofs of the basic theorems on queues and the starting point of the story for Poisson processes. Bauerschmidt's notes.

Part 3 Advanced Probability

Would really like to get hold of an interesting almost current concrete research problem to give direction to the theory. Martingale convergence theorems and basic secret examples of stochastic integrals and Feynman-Kac.

Part 3 Stochastic Calculus

Would really like to get hold of an interesting almost current concrete research problem to give direction to the theory.

Number Theory

Greg is taking this in the spring semester though I probably can't fit it into my schedule and the registration is already full last I checked anyways. My main interest in this is to find motivation for commutative algebra from a source other than algebraic geometry, and also to see what the algebraic number theorists have to say about solving diophantine equations, which seems to be a basic and interesting problem, as introduced by Shafarevich.

Not sure what books or notes to start with. Could look into the Cambridge notes and the standard textbook for Greg's course.

Algebraic Geometry

Want to understand (moduli spaces of) J-holomorphic curves in symplectic geometry. This was invented by Gromov and apparently inspired by pre-existing theory of holomorphic curves in algebraic geometry. Sounds to me it is basically about counting how many lines or curves of a particular degree there are in an algebraic variety. Like "27 lines on a cubic hypersurface" kind of thing, which is of course the classical starting point of the mirror symmetry business.

So I should learn about (moduli spaces of) algebraic curves. And probably Riemann surfaces and compare and contrast. "Curves and their Jacobians" seems to be a thing. Riemann surfaces version is in the Rick Miranda book.

Prerequisite for reading about curves and their Jacobians is to review and strengthen general basic algebraic geometry e.g. from Shafarevich, Kempf, Mumford. Would like to also take the opportunity to learn about sheaves, sheaf cohomology, and schemes. Mumford's Red Book is probably the desired "endpoint", though I am not sure how it compares vs Hartshorne. Shafarevich is very nice and concrete but does not talk about sheaves or schemes. Kempf would be a useful stepping stone.

So probably review Shafarevich for concrete elementary examples, read Kempf for the sheaves, read Mumford for the schemes, then read about moduli spaces of curves and Jacobians of curves. I suddenly recall something about Hilbert polynomial or Hilbert scheme which I should also look into. Of course on the Riemann surfaces side there is also the thing about period integrals and Hodge theory that I have wanted to look into for the longest time ever.

Symplectic Geometry

Want to understand framework and applications and computations of Morse theory and Floer homology, in particular J-holomorphic curves for the Floer homology and also more generally in symplectic geometry other than Floer homology. As always, Audin & Damian and the original papers of Floer, but also Abouzaid for the symplectic homology and cohomology and isomorphisms with string or loop homology for the non-compact manifolds.

Other basic concepts: Liouville manifolds, contact manifolds, Legendrian submanifolds, conical ends, Lagrangian torus fibrations over an affine base, almost complex structures which of course reminds one of Kähler manifolds and Calabi-Yau manifolds, the canonical and anticanonical bundles.

3-Manifolds

The basic state of knowledge of 3-manifolds. Required for writing the paper with Nathan when the time comes again. Classification of curves on surfaces and the proof that the algorithm works via flip moves or arcslides.

Differential Geometry

Origins of the subject leading to theories of Riemannian metrics, curvature, and connections. Spivak of course, but also PMH Wilson, which might remind me of some questions regarding Lie groups and hence topological and algebraic groups.

Algebraic Topology

Fundamental class, characteristic classes, computations of homology and cohomology using simplicial and other theories, isomorphisms between these theories with a view towards computations. Spin manifolds.

General brainteasers and problem solving exercises

Literally to prepare for interviews and online assessments. Would like to solve and create my own collection. Start from quant interview guide books and 1A Numbers & Sets but also look into olympiad problems and Putnam and other competitions.

Day 1, 27th December 2024, Friday

Q1

If $p \in Z(I(X))$ then for every polynomial $f \in k[x_1, \dots, x_n]$ such that $f = 0$ on X, have $p \in Z(f)$. So for every Zariski closed subset of $\mathbb{A}^n$ containing X, say defined by (the vanishing of) polynomials $f_1, \dots, f_m$, have $p \in Z(f_1, \dots, f_m)$, i.e. p is in every Zariski closed subset of $\mathbb{A}^n$ containing X.

Conversely, suppose $p \in \mathbb{A}^n$ is contained in every Zariski closed subset of $\mathbb{A}^n$ containing X. Then for any $f \in I(X)$, since $f = 0$ on X, have that X is contained in the Zariski closed subset $Z(f)$. Hence $p \in Z(f)$. This shows that $f(p) = 0$ for every $f \in I(X)$ so $p \in Z(I(X))$.

Q2

Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety and $U \subseteq X$ a non-empty open set. Let Y be a closed subset of $\mathbb{A}^n$ containing U, say $Y = Z(f_1, \dots, f_m)$ for some polynomials $f_1, \dots, f_m$. Suppose there is some $p \in X$ that is not contained in Y, so $f_i(p) \neq 0$ for some $i \in \{1, \dots, m\}$, wlog say $f_1(p) \neq 0$. In particular, $p \notin U$ since $U \subseteq Y$. Write $X_1 = X - U$ and $X_2 = X \cap Z(f_1)$. Clearly these are both closed subsets of X, and since $U \subseteq Z(f_1)$, have that $X = X_1 \cup X_2$. But X_1 is not the whole of X since U is nonempty, and X_2 is not the whole of X since $p \notin X_2$. This gives X as a union of two proper closed subsets, contradicting the irreducibility of X. Thus it is impossible for there to be a closed subset $Y \supseteq U$ but with $p \in X$ and $p \notin Y$. That is, every closed subset of $\mathbb{A}^n$ containing U must contain all of X, i.e. U is dense in X.

Suppose U is not irreducible, so we can write $U = U_1 \cup U_2$ where U_1, U_2 are proper closed subsets of U. Let $X_1, X_2 \subseteq X$ be closed subsets of X s.t. $U_1 = U \cap X_1$ and $U_2 = U \cap X_2$. Clearly $X_1 \neq X$ since $U_1 \neq U$, and similarly $X_2 \neq X$. But $U = U_1 \cup U_2 = (U \cap X_1) \cup (U \cap X_2) = U \cap (X_1 \cup X_2)$ so $X_1 \cup X_2$ is a closed subset of X containing all of U. If $X_1 \cup X_2 = X$, then we would have obtained X as a union of two proper closed subsets, contradicting the irreducibility of X. But if $X_1 \cup X_2$ is not all of X, then writing $X_3 = X - U$ we definitely have that $X = (X_1 \cup X_2) \cup X_3$ as a union of two proper closed subsets, again contradicting the irreducibility of X. Hence it is impossible to write U as a union of proper closed subsets, i.e. U is irreducible.

Suppose X is an irreducible affine variety with at least 2 distinct points p and q. If X is Hausdorff, then there are disjoint open subsets U and V of X containing p and q respectively. Write $Y = X - U$ and $Z = X - V$. These are proper closed subsets of X since $p \notin Y$ and $q \notin Z$. Furthermore, since U and V are disjoint, have $V \subseteq Y$ and $U \subseteq Z$ so $Y \cup Z = X$. This gives X as a union of proper closed subsets, contradicting the irreducibility of X. Hence it is impossible for X to have two distinct points p and q and be Hausdorff.

Q3

Let $X \supseteq X_1 \supseteq X_2 \supseteq X_3 \supseteq \dots$ be a descending chain of closed subsets in an affine variety $X \subseteq \mathbb{A}^n$. Then $I(X_1) \subseteq I(X_2) \subseteq I(X_3) \subseteq \dots$ is an increasing chain of ideals in $k[x_1, \dots, x_n]$, which is a Noetherian ring, so there is some N s.t. $I(X_n) = I(X_N)$ for all $n \geq N$. Since the X_i are closed subsets, $Z(I(X_i)) = X_i$ by Q1, so $X_n = Z(I(X_n)) = Z(I(X_N)) = X_N$ for all $n \geq N$, i.e. the descending chain of closed subsets is eventually constant.

Q4

The coordinate ring of Y is $k[x,y]/(xy-1)$ but the coordinate ring of $\mathbb{A}^1$ is $k[z]$. An isomorphism between Y and $\mathbb{A}^1$ induces an isomorphism between their coordinate rings. In particular, $x, y \in k[x,y]/(xy-1)$ will be sent under this isomorphism to polynomials $p(z), q(z) \in k[z]$ such that $p(z)q(z) = 1$. But this can only happen if $p(z)$ and $q(z)$ are constants, i.e. if $\mathbb{A}^1$ is mapped to a single point in Y by the isomorphism, which is impossible (because k is algebraically closed so both Y and $\mathbb{A}^1$ have infinitely many points).

As argued above, any morphism $\mathbb{A}^1 \to Y$ induces a pullback of the coordinate functions x and y on Y, sending them to polynomial functions $p(z)$ and $q(z)$ on X, which must satisfy $p(z)q(z) = 1$, which is only possible if x and y are constants. So the only morphism $\mathbb{A}^1 \to Y$ is the constant map.

Regarding morphisms $Y \to \mathbb{A}^1$, there are the two obvious projections onto the coordinate axes $(x,y) \mapsto x$ and $(x,y) \mapsto y$. More generally, we can project onto any straight line through the origin, $(x,y) \mapsto ax+by$ for some constants $a, b \in k$, and then compose with any morphism $\mathbb{A}^1 \to \mathbb{A}^1$, which is simply a polynomial function in one variable, to obtain $(x,y) \mapsto p(ax+by)$ for any $p(z) \in k[z]$.