Assignment 7

Modern Theory of Markov Chains

Due: 28.05.2013

1 (Couplings and monotonicity)**.** A coupling of two or more random variables is nothing but a simultaneous reconstruction of (copies of) the variables in the same probability model. For example, let X and Y are Bernoulli random variables each taking values $\mathtt{0}$ and $\mathtt{1}$ with probability $1/2$. We can construct a coupling $(\tilde{X}, \tilde{Y})$ of X and Y by generating a uniform random variable Z over the set $\{\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}\}$ and defining $\tilde{X}$ as the first bit of Z and $\tilde{Y}$ as the second bit of Z. Another equally valid coupling is obtained similarly if Z is chosen non-uniformly according to the distribution p with $p(\mathtt{00}) = p(\mathtt{11}) = 1/8$ and $p(\mathtt{01}) = p(\mathtt{10}) = 3/8$.

a) Let X and Y be non-uniform Bernoulli random variables such that $\mathbf{P}(X = \mathtt{1}) \leq \mathbf{P}(Y = \mathtt{1})$. Find a coupling $(\tilde{X}, \tilde{Y})$ of X and Y such that $\mathbf{P}(\tilde{X} \leq \tilde{Y}) = 1$.

Couplings are often helpful to give rigorous proofs of intuitively plausible facts.

b) (Percolation probability) Consider a connected graph $G = (V, E)$ and two distinct vertices $u, v \in V$. Let H be a random subgraph obtained from G by flipping a biased coin for each edge to decide whether the edge is kept or removed. Suppose therefore that each edge is kept (independently of the others) with probability $0 < p < 1$ and is removed otherwise. Let $\{u \stackrel{H}{\longleftrightarrow} v\}$ denote the event that u and v are connected in H. Prove that $\mathbf{P}(u \stackrel{H}{\longleftrightarrow} v)$ is an non-decreasing function of the parameter p.

c) (Galton-Watson) Prove that the probability of survival in the Galton-Watson model with parameter p (i.e., a population of bacteria where each bacterium is divided to two new bacteria with probability p or dies with probability $1-p$) is monotone with respect to p.

Can you imagine a simple proof of the last two statements without using the idea of coupling?

(BONUS) **2** (Total variation distance)**.** Verify that the following definitions of the total variation distance between two probability distributions μ and ν on a finite set S are equivalent. What is the intuitive meaning of each of these definitions?

a) $\|\mu - \nu\| \triangleq \sup\{|\mu(E) - \nu(E)| : \ E \subseteq S\}$.

b) $\|\mu - \nu\| \triangleq \frac{1}{2}\sum_{a\in S} |\mu(a) - \nu(a)|$.

c) $\|\mu - \nu\| \triangleq \sup\{|\mu(f) - \nu(f)| : \ f : S \to \mathbb{R} \text{ and } 0 < \|f\| \leq 1\}$,
where $\|f\| \triangleq \sup_{a\in S} |f(a)|$ is the supremum norm.

The total variation distance is the metric generated by the total variation norm for signed distributions, hence the notation $\|\mu - \nu\|$ rather that $d_{\mathrm{TV}}(\mu, \nu)$. For an infinite space S, similar definitions can be given and will again be equivalent, although we should impose measurability conditions on E and f.

3 (Coupling inequality)**.** Let S be a finite set and μ and ν two probability distributions on S. By a coupling of μ and ν we mean a pair (X, Y) of random variables on S such that X is distributed according to μ and Y is distributed according to ν.

a) Show that every coupling (X, Y) of μ and ν satisfies the inequality $\|\mu - \nu\| \leq \mathbf{P}(X \neq Y)$.

A coupling (X, Y) of μ and ν satisfying $\|\mu - \nu\| = \mathbf{P}(X \neq Y)$ is called an *optimal* coupling. Note that among all couplings of μ and ν, an optimal coupling has the highest probability of agreement $\mathbf{P}(X = Y)$. Do optimal couplings exist for every two distributions?

b) Let $S = \{a, b, c\}$ and μ and ν be the distributions with

$$\mu(a) = \mu(b) = \mu(c) = \frac{1}{3}$$
$$\nu(a) = \frac{1}{4}, \quad \nu(b) = \frac{1}{2}, \quad \nu(c) = \frac{1}{4}.$$

Find an optimal coupling of μ and ν.

(BONUS) c) Prove that any two probability distributions μ and ν on a finite set S have an optimal coupling.

Combining (a) and (c) we have yet another interpretation of the total variation distance:

$$\|\mu - \nu\| = \inf\{\mathbf{P}(X \neq Y) : (X, Y) \text{ a coupling of } \mu \text{ and } \nu\}.$$

4 (Coupling of Markov chains)**.** A coupling of two Markov chains $(X_t)_t$ and $(Y_t)_t$ is a simultaneous construction of the random variables $\tilde{X}_0, \tilde{X}_1, \ldots$ and $\tilde{Y}_0, \tilde{Y}_1, \ldots$ in the same probability model in such a way that the variables $(\tilde{X}_t)_t$ have the same joint distributions as $(X_t)_t$ and the variables $(\tilde{Y}_t)_t$ have the same joint distributions as $(Y_t)_t$.

a) Let $(\tilde{X}_t, \tilde{Y}_t)_t$ be a coupling of two copies of a Markov chain with transition probabilities K and initial distributions μ and ν, respectively. Let T be a random variable on $\mathbb{N} \cup \{\infty\}$ with the property that for all $t \geq T$, we have $\tilde{X}_t = \tilde{Y}_t$. Verify that $\|\mu K^t - \nu K^t\| \leq \mathbf{P}(T \geq t)$. As we have seen, this is the basis of coupling arguments used to bound the speed of convergence of Markov chains.

In practice, the couplings used in the study of Markov chains often have the special property that the sequence $(\tilde{X}_t, \tilde{Y}_t)_t$ itself is a (time-homogeneous) Markov chain. Such a coupling is called a *Markov* coupling and can be constructed by introducing a coupling $\tilde{K}((a, b), \cdot)$ of the distributions $K(a, \cdot)$ and $K(b, \cdot)$ for every two states a and b.

b) Give an example of a coupling of two copies of a Markov chain that is *not* a Markov coupling.

c) For Markov couplings, it is convenient that the two chains remain together as soon as they meet. Express this condition in terms of the transition matrix $\tilde{K}$. Observe that for such a coupling, the random variable $T \triangleq \inf\{t \geq 0 : \tilde{X}_t = \tilde{Y}_t\}$ satisfies the assumption of part (a).

5 (Forgetful chain)**.** Let K be the transition matrix of a simple random walk on $\mathbb{N} = \{0, 1, \ldots\}$ with reflection at the origin, that is,

$$K(i, j) \triangleq \begin{cases} \frac{1}{2} & \text{if } i > 0 \text{ and } |j - i| = 1, \\ 1 & \text{if } i = 0 \text{ and } j = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let us the denote the (standard) lazy version of K by $\tilde{K}$, that is, $\tilde{K} = {}^1\!/\!_2(I + K)$, where I is the identity matrix on $\mathbb{N}$.

Recall that $\tilde{K}$ has no stationary distribution (why?), and hence the convergence theorem does not apply. Nevertheless, a particle doing a random walk according to $\tilde{K}$ asymptotically "forgets" its initial position in the following sense: for any two starting points $a, b \in \mathbb{N}$, the total variation distance of the distributions $K^t(a, \cdot)$ and $K^t(b, \cdot)$ approaches 0 as $t \to \infty$. Can you prove this forgetfulness?

Hint: Construct a Markov coupling of two copies of the walk with starting points a and b as follows. Before the two particles meet, at each step, flip a coin. If it comes heads, move the first particle according to K. Otherwise, move the second particle according to K. Once the two walks meet in the same position, let them continue their walk together. Use the fact that position 0 is recurrent (why?).