Statistics 251/551 spring 2013
2013: Solutions to sheet 3

If you are not able to solve a part of a problem, you can still get credit for later parts: Just assume the truth of what you were unable to prove in the earlier part.

[1] *(10 points) Find the transition probability matrix Q^* for the bivariate chain $Z_n^* = (X_n Y_n^*)$ that appeared in the coupling proof of the BLT. It would be a good idea to check that*

$$\sum_{z_1 \in \mathcal{S}\times\mathcal{S}} Q^*(z_0, z_1) = 1 \qquad \text{for each } z_0.$$

By definition, $T = \min\{n : X_n = Y_n\}$ and

$$Y_n^* = \begin{cases} Y_n & \text{if } n < T \\ Y_n = X_n & \text{if } n = T \\ X_n & \text{if } n > T \end{cases}.$$

I asserted that $Z_n^* := (X_n, Y_n^*)$ is a Markov chain with stationary transition probabilities. That is, for all z_0 and z_1 in $\mathcal{S}\times\mathcal{S}$ there should exist nonnegative (nonrandom) numbers $Q^*(z_0, z_1)$ for which $\sum_{z_1 \in \mathcal{S}\times\mathcal{S}} Q^*(z_0, z_1) = 1$ and

$$\mathbb{P}\{Z_{n+1}^* = z_1 \mid Z_n^* = z_0, \text{ past info}\} = Q^*(z_0, z_1) \qquad \text{for all } z_0 \text{ and } z_1 \text{ in } \mathcal{S}\times\mathcal{S}.$$

Many of you gave values for $Q^*(z_0, z_1)$ than seemed to depend on the random variable T. That is, you made Q^* random, which is illegal. Obviously T has to get into the story somehow but the result must not be random.

Here is how to use T properly. Suppose $z_0 = (i_0, j_0)$ and $z_1 = (i_1, j_1)$.

Consider first the case where $i_0 \neq j_0$. If $X_n = i_0$ and $Y_n^* = j_0$ then $Y_n^* \neq X_n$. The events $\{Z_n^* = z_0\}$ and $\{X_n = i_0, Y_n = j_0, T > n\}$ are the same. If we condition on that event then we must have $Y_{n+1}^* = Y_{n+1}$. Thus

$$\begin{aligned}
&\mathbb{P}\{Z_{n+1}^* = z_1 \mid Z_n^* = z_0, \text{ past info}\} \\
&\quad = \mathbb{P}\{X_{n+1} = i_1, Y_{n+1} = j_1 \mid X_n = i_0, Y_n = j_0, T > n, \text{ past info}\} \\
&\quad = \mathbb{P}\{X_{n+1} = i_1, Y_{n+1} = j_1 \mid X_n = i_0, Y_n = j_0\} \\
&\quad = P(i_0, i_1) P(j_0, j_1).
\end{aligned}$$

In the third line, the past info and the event $\{T > n\}$ (which tells us something about X_m and Y_m for $m \leq n$) become irrelevant if we also condition on $X_n = i_0$ and $Y_n = j_0$. That's the Markov property. The factorization in the last line then comes from the independence of the X- and Y-chains.

Now consider the case where $i_0 = j_0$. If $Y_n^* = j_0 = i_0 = X_n$ then we must have $T \leq n$; the X- and Y-chains must have met at time n or earlier. The two events $\{Z_n^* = z_0\}$ and $\{X_n = i_0, T \leq n\}$ are the same. Thus

$$\begin{aligned}
&\mathbb{P}\{Z_{n+1}^* = z_1 \mid Z_n^* = z_0, \text{ past info}\} \\
&\quad = \mathbb{P}\{X_{n+1} = i_1, X_{n+1} = j_1 \mid X_n = i_0, T \leq n, \text{ past info}\} \\
&\quad = \begin{cases} P(i_0, i_1) & \text{if } i_1 = j_1 \\ 0 & \text{if } i_1 \neq j_1 \end{cases}
\end{aligned}$$

Again the past info and the event $\{T \le n\}$ become irrelevant once we condition on $X_n = i_0$.

In summary, the Z^*-chain has transition probabilities

$$Q^*\left[(i_0, j_0), (i_1, j_1)\right] = \begin{cases} P(i_0, i_1)P(j_0, j_1) & \text{if } i_0 \ne j_0 \\ P(i_0, i_1) & \text{if } i_0 = j_0 \text{ and } i_1 = j_1 \\ 0 & \text{otherwise} \end{cases}$$

The stuff about T caused the differences between the $i_0 = j_0$ and $i_0 \ne j_0$ cases but T itself appears nowhere is the expression for Q^*.

[2] *(15 points) Chang Exercise 1.17.*

(a) (Why a MC with state space $\mathcal{S}$?) The update step ensures that the row and column sums do not change. Each table belongs to $\mathcal{S}$. The Markov property comes from the fact that $\mathbb{P}\{X_{n+1} = \tau \mid X_n = \sigma, \text{ past info}\}$ is determined by the same probabilty mechansim, which depends only on σ, each time X_n is in state σ.

(c) (Aperiodic?) In what follows I draw pictures as if $i_1 = j_1 = 1$ and $i_2 = j_2 = 2$ and focus only on changes that affect that top left corner of the table. (The values shown as dots are unaffected.) There are $\binom{4}{2}^2 = 36$ ways to actually choose i_1, i_2, j_1, j_2.

$$\sigma = \begin{array}{|c|c|c|c|c}
\text{a} & \text{b} & . & . & 220 \\
\text{c} & \text{d} & . & . & 215 \\ \hline
. & . & . & . & 93 \\
. & . & . & . & 64 \\ \hline
108 & 286 & 71 & 127 &
\end{array}$$

$$\sigma' = \begin{array}{|c|c|c|c|c}
a+1 & b-1 & . & . & 220 \\
c-1 & d+1 & . & . & 215 \\ \hline
. & . & . & . & 93 \\
. & . & . & . & 64 \\ \hline
108 & 286 & 71 & 127 &
\end{array}
\qquad
\sigma'' = \begin{array}{|c|c|c|c|c}
a-1 & b+1 & . & . & 220 \\
c+1 & d-1 & . & . & 215 \\ \hline
. & . & . & . & 93 \\
. & . & . & . & 64 \\ \hline
108 & 286 & 71 & 127 &
\end{array}$$

If $\min(a, b, c, d) > 0$ then $P(\sigma, \sigma') = 1/72 = P(\sigma, \sigma'')$. If $\min(a, b, c, d) = 0$ the probabilities are a little different. For example, if $a = 0 < b, c, d$ then $P(\sigma, \sigma) = 1/72 = P(\sigma, \sigma')$. Tables in $\mathcal{S}$ that contain at least one zero (of which there are many) have period 1. If the chain is irreducible then it has period 1.

(d) (Stationary distribution uniform?) The transition probabilities are symmetric: $P(\sigma, \tau) = P(\tau, \sigma)$. (Time reversibility.) If $\sigma \ne \tau$ and the chain goes from σ to τ by choosing i_1, i_2, j_1, j_2 and tossing a head then the chain goes from τ to σ by choosing i_1, i_2, j_1, j_2 and tossing a tail.

If $\pi_\sigma = 1/\#\mathcal{S}$ for all σ in $\mathcal{S}$ then

$$\sum\nolimits_{\sigma\in\mathcal{S}} \pi_\sigma P(\sigma,\tau) = \sum\nolimits_{\sigma\in\mathcal{S}} \pi_\tau P(\tau,\sigma) = \pi_\tau$$

That is, π is the stationary distribution.

(b) (Irreducible?) One way to proceed is to work with the distance between tables defined by $d(\sigma,\tau) := \sum_{i,j} |\sigma[i,j] - \tau[i,j]|$. It suffices to show that if $\sigma \neq \tau$ then there is a always a legal step from σ to some σ' for which $d(\sigma',\tau) < d(\sigma,\tau)$. By a finite number of such steps we could build a legal path between σ and τ.

Locate a suitable place for taking the step by looking at the signs in the table of differences $D = \sigma - \tau$. The row and column sums of D are all zero. Also D must contain at least one strictly negative number. For concreteness suppose $D[1,1] < 0$. To make the sum across the first row equal to zero there must exist some column j for which $D[1,j] > 0$. Similarly, there must exist some row i for which $D[i,1] > 0$. We have no control over $D[i,j]$. The pattern of signs in D shows that

$$\begin{array}{rl} \sigma[1,1] < \tau[1,1] & \quad \sigma[1,j] > \tau[1,j] \\ \sigma[i,1] > \tau[i,1] & \quad ?? \end{array}$$

In consequence, both $\sigma[i,1]$ and $\sigma[1,j]$ are strictly positive. The table σ' that agrees with σ except for

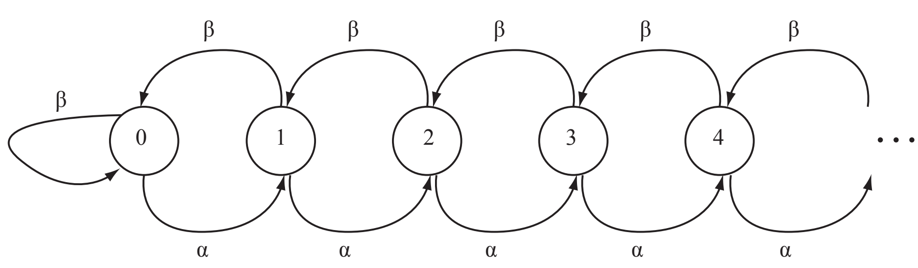
$$\begin{array}{rl} \sigma'[1,1] = \sigma[1,1] + 1 & \quad \sigma'[1,j] = \sigma[1,j] - 1 \\ \sigma'[i,1] = \sigma[i,1] - 1 & \quad \sigma'[i,j] + 1 \end{array}$$

contains no strictly negative values (so it belongs to $\mathcal{S}$) and

$$\begin{aligned} |\sigma'[1,1] - \tau[1,1]| &= |\sigma[1,1] - \tau[1,1]| - 1 \\ |\sigma'[1,j] - \tau[1,j]| &= |\sigma[1,j] - \tau[1,j]| - 1 \\ |\sigma'[i,1] - \tau[i,1]| &= |\sigma[i,1] - \tau[i,1]| - 1 \\ |\sigma'[i,j] - \tau[i,j]| &= |1 + \sigma[i,j] - \tau[i,j]| \leq 1 + |\sigma[i,j] - \tau[i,j]|. \end{aligned}$$

We have made three improvements at a possible cost of at most one in the $[i,j]$ position. Thus $d(\sigma',\tau) \leq d(\sigma,\tau) - 2$.

[3] Consider once again the one-sided random walk described on Wednesday 23 January (and on the onesideRW.pdf handout). Suppose $\beta > 1/2$. You know that the chain has a stationary probability distribution. The chain is recurrent. In fact (see bonus question) it is positive recurrent. Find the value $\tau := \mathbb{E}_1 T_0$ by these steps.



(i) (10 points) Explain why τ must be finite if the chain is positive recurrent.

(ii) (10 points) Show that $\mathbb{E}_k T_0 = k\tau$ for each $k \geq 1$.

(iii) (10 points) Set up an equation for τ by conditioning on the first step, then solve. Your solution should tend to infinity as β decreases to $1/2$.

For (i): By conditioning on the first step we get

$$\infty > \mathbb{E}_0 T_0 = \beta(1+0) + \alpha(1 + \mathbb{E}_1 T_0)$$

Many of you forgot to add 1 for the first step.

For (ii), note that a path from k to 0 must first pass through $k-1$, then through $k-2$, ..., then through 1, then lead to zero. (In between visiting k_1 then visiting $k-2$ the chain might return to k.) The Markov property gives the representation

$$\mathbb{E}_k T_0 = \mathbb{E}_k T_{k-1} + \mathbb{E}_{k-1} T_{k-2} + \cdots + \mathbb{E}_1 T_0.$$

For $\mathbb{E}_k T_{k-1}$ notice that nothing to the left of state $k-1$ in the picture has any effect on the calculation. If we ignore those irrelevant parts we are left with a picture that looks just like the original one except that the labels are changed from $0, 1, \dots$ to $k-1, k, \dots$ (and it lacks a loop from $k-1$ back to $k-1$). The task of getting from k to $k-1$ is probabilistically the same as the task of getting from 1 to 0. Hence $\mathbb{E}_k T_{k-1} = \mathbb{E}_1 T_0 = \tau$. And so on.

For (iii), condition on $F = \{\text{first step goes to } 0\}$ and $F^c = \{\text{first step goes to } 2\}$;

$$\tau = \mathbb{E}_1 T_0 = (\mathbb{P}_1 F)\,\mathbb{E}_1(T_0 \mid F) + (\mathbb{P}_1 F^c)\,\mathbb{E}_1(T_0 \mid F^c) = \beta(1) + \alpha(1 + \mathbb{E}_2 T_0) = 1 + 2\alpha\tau$$

whose only finite solution is $\tau = (1-2\alpha)^{-1}$.

[4] (bonus points) You know that an irreducible ***[aperiodic]*** *Markov chain that has a stationary probability distribution $\pi = \{\pi_i : i \in \mathcal{S}\}$ must be recurrent. Use the BLT to show that such a chain must actually be positively recurrent. Here are some ideas that might help.*

For some state i suppose $\mathbb{E}_i T_i = +\infty$.

(i) Let $V_N = \sum_{n \leq N} \mathbb{I}\{X_n = i\}$ denote the number of visits to state i in the first N steps. Use the BLT to show that $\mathbb{E}_i V_N / N \to \pi_i$ as $N \to \infty$.

(ii) For each positive integer M, show that $V_N \geq M$ iff $\sum_{n \leq M} T_i^{(n)} \leq N$, where the $T_i^{(n)}$'s are the successive cycle times.

(iii) With $\mathbb{P}_i$ probability one, $\sum_{n \leq M} T_i^{(n)}/M \to \infty$. (Why?) Deduce, for each $\epsilon > 0$, that

$$\mathbb{P}_i\{V_N/N > \epsilon\} \to 0 \qquad \text{as } N \to \infty.$$

(iv) Deduce that $\mathbb{E}_i(V_N/N) < 2\epsilon$ for N large enough.

(v) What does that tell you about π_i?

(vi) In fact, is it possible to have $\pi_j = 0$ for at least one j in $\mathcal{S}$?

My apologies for omitting the word *aperiodic* from the original statement of the Problem. None of you seemed to be bothered by the omission. I would be interested to hear from any of you who can correctly describe (with proofs) what happens if the chain has a stationary probability distribution π, is irreducible, and has period 2.

For (i):

$$\mathbb{E}_i V_N = \sum\nolimits_{1\le n\le N} \mathbb{E}_i \mathbb{I}\{X_n = i\} = \sum\nolimits_{1\le n\le N} \mathbb{P}_i\{X_n = i\}.$$

By the BLT, $\mathbb{P}_i\{X_n = i\} \to \pi$ as $n \to \infty$. Then use the Analysis fact that if $a_n \to a$ then $N^{-1}\sum_{1\le n\le N} a_n \to a$.

Remark. See *http://en.wikipedia.org/wiki/Cesàro_summation*. The proof is easy. For each $\epsilon > 0$ there exists an n_ϵ such that $|a_n - a| < \epsilon$ for all $n \ge n_\epsilon$. For $N > n_\epsilon$,

$$\left| a - N^{-1}\sum\nolimits_{1\le n\le N} a_n \right| \le N^{-1}\sum\nolimits_{1\le n\le n_\epsilon} |a_n - a| + \frac{N - n_\epsilon}{N}\epsilon$$

On the right-hand side, the first sum is fixed while N goes to infinity and the second term is less than ϵ.

Many of you seemed to assert that $\mathbb{P}_i\{X_n = i\} = \pi_i$ for all n (which is not true), or started a calculation with $\mathbb{P}_\pi$ with a mysterious change to $\mathbb{P}_i$ in subsequent lines.

For (ii), I intended the calculation to refer to chains starting from state i. The asserted inequality reflects the fact that the Mth return to state i occurs at time $S_M := \sum_{n\le M} T_i^{(n)}$, the end of the Mth cycle.

For (iii) I was expecting something like: for each $\epsilon > 0$

$$\mathbb{P}_i\{\sum\nolimits_{n\le M} T_i^{(n)}/M \le 2/\epsilon\} \to 0 \qquad \text{as } M \to \infty.$$

Take M as the largest integer for which $M \le N\epsilon$. Then

$$\begin{aligned}\mathbb{P}_i\{V_N > N\epsilon\} &\le \mathbb{P}_i\{V_N \ge M\} \\ &= \mathbb{P}_i\{\sum\nolimits_{n\le M} T_i^{(n)}/M \le N/M\} \le \mathbb{P}_i\{\sum\nolimits_{n\le M} T_i^{(n)}/M \le 2/\epsilon\},\end{aligned}$$

which tends to zero as N tends to infinity.

Some of you asserted that $S_{V_N} = N$, which is not true. (Consider the case $T_i^{(1)} = T_i^{(2)} = T_i^{(3)} = 4$ and $N = 9$, for which $V_N = 2$ and $T_i^{(1)} + T_i^{(2)} = 8$.) However it is true that $S_{V_N} \le N < S_{1+V_N}$. For each realization of the chain along which $S_n/n \to \infty$ as $n \to \infty$ we have

$$N/V_N \ge S_{V_N}/V_N \to \infty.$$

Thus $V_N/N \to 0$ with $\mathbb{P}_i$ probability one.

Those of you who have studied some advanced probability would have recognized the argument leading from

$$\mathbb{P}_i\{V_N/N \to 0\} = 1 \qquad \text{to} \qquad \mathbb{P}_i\{V_N/N > \epsilon\} \to 0 \text{ for each } \epsilon > 0,$$

or from

$$\mathbb{P}_i\{S_n/n \to \infty\} = 1 \qquad \text{to} \qquad \mathbb{P}_i\{S_n/n \le C\} \to 0 \text{ for each finite } C,$$

as an example of almost sure convergence implying convergence in probability. I was not expecting any of you to know any advanced probability; I accepted any reasonable attempt to explain.

For (iv):

$$\mathbb{E}_i(V_N/N) \le \mathbb{E}_i\left(\epsilon\mathbb{I}\{V_N \le N\epsilon\} + \mathbb{I}\{V_N > N\epsilon\}\right) \le \epsilon + \mathbb{P}\{V_N > N\epsilon\}.$$

For (v): Parts (i) and (iv) togther imply that $\pi_i < 2\epsilon$ for each $\epsilon > 0$. That is, π_i must be zero.

For (vi): By HW2.2, if $\mathbb{E}_i T_i = \infty$ for one $i \in \mathcal{S}$ then $\mathbb{E}_j T_j = \infty$ for all $j \in \mathcal{S}$, which would imply $\pi_j = 0$ for all j. That is, π could not be a probability distribution. The only alternative is that $\mathbb{E}_j T_j < \infty$ and $\pi_j = 1/\mathbb{E}_j T_j > 0$ for all j.

Some of you gave a more direct argument. For each $j \in \mathcal{S}$ and each $n \in \mathbb{N}$,

$$\pi_j = \sum\nolimits_{i \in \mathcal{S}} \pi_i P^n(i,j).$$

There must exist at least one i_0 for which $\pi_{i_0} > 0$, because $\sum_{i \in \mathcal{S}} = 1$. If $\pi_j = 0$ then we would have $P^n(i_0, j) = 0$ for all n, which would violate the assumption that $i_0 \rightsquigarrow j$.

Some of you were confused by the fact, for a nonnegative random variable T, that $\mathbb{P}\{T < \infty\} = 1$ is compatible with $\mathbb{E}T = \infty$.