Statistics 251b/551b, spring 2009
Solutions for test 1

- *You should prepare solutions to the following questions without help or hints from anybody else. In particular, if you have been working in a group you must suspend the group arrangement for the test; you must not discuss the questions with your group buddy.*
- *If you have questions of interpretation, or of clarification of the meaning of a question, ask David Pollard.*
- *Make sure you explain your calculations and notation.*
- *Each part of each question is worth 5 points.*
- *As usual, even if you are unable to solve one part of a question you may still use the result for the following parts.*

*[1]* *For each of the following questions, $\mathcal{S}$ denotes the set $\{0, 1, 2, 3, \dots\} = \{0\} \cup \mathbb{N}$.*

*(i) Give an example of an irreducible Markov chain with state space $\mathcal{S}$ for which there exists a stationary measure but no stationary probability distibution.*

Many of you used the the chain discussed in lecture 3, with $\alpha = 1/2$. That is, $P(0,0) = P(0,1) = 1/2$ and $P(i, i-1) = 1/2 = P(i, i+1)$ for $i \geq 1$. As shown in class, the only solution to the equations $\lambda_j = \sum_{i \in \mathcal{S}} \lambda_i P(i,j)$ for $j \in \mathcal{S}$ is $\lambda_i = \lambda_0$ for all $i$. Taking $\lambda_0 = 1$ we get a stationary measure but there is no way to choose $\lambda_0$ to make $\sum_{i \in \mathcal{S}} \lambda_i = 1$.

*(ii) Give an example of a Markov chain with state space $\mathcal{S}$ that has at least two distinct stationary probability distibutions.*

The simplest example makes every state absorbing, so that every probability measure on $\mathcal{S}$ is a stationary probability distibution.

*(iii) Give an example of a martingale taking values in $\mathcal{S}$ that is not a Markov chain.*

The challenge is to define conditional distributions

$$p_n(i \mid i_0, \dots, i_{n-1}) = \mathbb{P}\{X_n = i \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}$$

so that

(i) $p_n$ depends on more than $i_{n-1}$ alone

(ii) $\sum_i i p_n(i \mid i_0, \dots, i_{n-1}) = i_{n-1}$ for all $i_{n-1}$.

Requirement (ii) forces us to take $p_n(0 \mid i_0, \dots, i_{n-1}) = 1$ if $i_{n-1} = 0$.

The idea in the following example is to make all the conditional distributions depend on $X_0$. Start with $\mathbb{P}\{X_0 = 1\} = 1/2 = \mathbb{P}\{X_0 = 2\}$. Let the process make jumps $\pm X_0$ each with conditional probability $1/2$. Of course, the case $i_{n-1} = 1$ needs special treatment to stop the process from jumping outside $\mathcal{S}$. Take

$$p_n(2 \mid i_0, \ldots, i_{n-1}) = p_n(0 \mid i_0, \ldots, i_{n-1}) = 1/2 \qquad \text{if } i_{n-1} = 1.$$

and, if $i_{n-1} \geq 2$ and $i_0 = k$, take

$$p_n(i_{n-1} + k \mid i_0, \ldots, i_{n-1}) = p_n(i_{n-1} - k \mid i_0, \ldots, i_{n-1}) = 1/2.$$

*(iv) Give an example of an irreducible, aperiodic Markov chain with state space $\mathcal{S}$ for which $\mathbb{P}_i\{X_n = i\} = 0$ for $1 \leq n \leq 100$ for every state $i$.*

I don't like having to worry about $101$ versus $100$. So take

$$P(0, j) = \begin{cases} C2^{-j} & \text{if } j \geq 1000 \\ 0 & \text{otherwise,} \end{cases}$$

where $C$ is chosen to make $\sum_{j \in \mathcal{S}} P(0, j) = 1$. Then define $P(i, i-1) = 1$ for all $i \geq 1$. Clearly $0 \leftrightsquigarrow j$ for every $j > 0$, so the chain is irreducible. It takes at least $1001$ steps to get from $i$ back to $i$ for every $i$ in $\mathcal{S}$. (Consider the cases $i < 1000$ and $i \geq 1000$ separately.)

*[2] For a fixed positive integer $d$ let*

$$\mathcal{S} = \{(i_1, i_2) \in \mathbb{N} \times \mathbb{N} : 1 \leq i_1 \leq d \text{ and } 1 \leq i_2 \leq d\}$$

*denote the $d \times d$ lattice of points with integer coordinates running from $1$ to $d$. Let $X_n = (X_{n,1}, X_{n,2})$ be a random walk on $\mathcal{S}$: if $X_n = x \notin \{(1,1), (d,d)\}$ then $X_{n+1}$ has the equal probability of being at one of the neighbors of $x$. Make both $(1,1)$ and $(d,d)$ absorbing states. For example, if $x = (x_1, x_2)$ with $1 < x_i < d$ for $i = 1, 2$ then $P(x, y) = 1/4$ for each $y$ in the set*

$$\{(x_1, x_2 + 1),\ (x_1, x_2 - 1),\ (x_1 + 1, x_2),\ (x_1 - 1, x_2)\}$$

*For $x$ on the edges of the lattice, there are fewer neighbors and the transition probabilities will be slightly different.*

*Define $\tau = \inf\{n \in \mathbb{N} : X_n = (1,1)$ or $X_n = (d,d)\}$.*
*Define $B = \{\tau < \infty,\ X_\tau = (d,d)\}$.*
*Define $Z_n = X_{n,1} + X_{n,2}$ for each $n$.*

*(i) Write down the transition probabilities $P(x, y)$ when $x$ is on the edge of the lattice. (That is, at least one of $x_1$ and $x_2$ is equal to $1$ or $d$.)*

*(ii) Define $f(x) = x_1 + x_2$ for $x = (x_1, x_2) \in \mathcal{S}$. Show that $f$ is harmonic. That is, show that $\mathbb{E}\left(f(X_1) \mid X_0 = x\right) = f(x)$ for each $x$.*

Write $A$ for $\{(1,1), (d,d)\}$, the set of absorbing states.

As my original description of the chain as a random walk was imprecise, I collapsed the first two questions into: Find transition probabilities $P(x, y)$ when $x$ is on the boundary and $x \notin A$ to make the function $f$ harmonic.

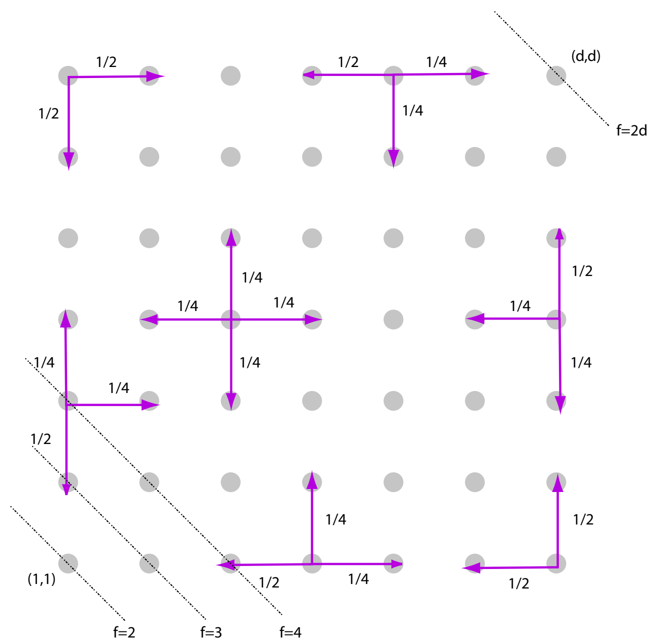
The assertion that $\mathbb{P}_x f(X_1) = f(x)$ when $x \in A$ holds for trivial reasons.

When $x = (x_1, x_2)$ is not on the boundary,

$$\begin{aligned}
&\mathbb{P}_x\{f(X_1) = f(x) + 1\} \\
&\quad = \mathbb{P}_x\{X_{n+1} = (x_1 + 1, x_2)\} + \mathbb{P}_x\{X_{n+1} = (x_1, x_2 + 1)\} \\
&\quad = 1/4 + 1/4 = 1/2
\end{aligned}$$

and, by a similar argument, $\mathbb{P}_x\{f(X_{n+1}) = f(x) - 1\} = 1/2$. It follows that

$$\mathbb{E}\left(f(X_{n+1}) \mid X_n = x)\right) = \tfrac{1}{2}\left(f(x) + 1\right) + \tfrac{1}{2}\left(f(x) - 1\right) = f(x).$$




To make $f$ harmonic everywhere, it suffices to have

$$\mathbb{P}_x\{f(X_1) = f(x) + 1\} = 1/2 = \mathbb{P}_x\{f(X_1) = f(x) - 1\}$$

for all boundry points $x \notin A$. The picture shows one possible choice for the transition probabilities. (In fact, we could also replace the $1/4$'s on the boundary by any other pair of nonnegative numbers that sum to $1/2$. The numbers could even change from one boundary point to another.)

*(iii) Explain why $\mathbb{P}_x\{\tau < \infty\} = 1$ for every $x$ in $\mathcal{S}$.*

For every $x$ in $\mathcal{S}$,

$$\mathbb{P}_x\{\tau \le 2d\} \ge \delta := 4^{-d},$$

because there is a path of at most $2d$ steps leading from each $x$ to an absorbing state. The usual chicken argument then gives

$$\begin{aligned}\mathbb{P}_x\{\tau > 2(k+1)d\} &= \sum\nolimits_{y \in \mathcal{S}\setminus A} \mathbb{P}_x\{\tau > 2kd, X_{2kd} = y\}\mathbb{P}_y\{\tau > 2d\} \\ &\le (1-\delta)\mathbb{P}_x\{\tau > 2kd\},\end{aligned}$$

leading to the bound $\mathbb{P}_x\{\tau > kd\} \le (1-\delta)^k$ for every $x$ and $k \in \mathbb{N}$. Question: Should I have considered the cases where $x \in A$ separately?

*(iv) Show that $Z_n$ is a martingale for each initial state $x$ of the $X_n$-chain.*

In class I showed, for a general Markov chain and a general harmonic function $f$, that $\mathbb{E}_x\left(f(X_{n+1}) - f(X_n) \mid D\right) = 0$ for every event $D$ of the form $\{X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\}$. That is, $f(X_n)$ is a martingale.

Some of you wanted to replace the conditioning event by an analogous event $F = \{Z_0 = z_0, \dots, Z_n = z_n\}$. For a formal argument you could have rewritten $F$ as a disjoint union of events of the form $D(x_{0,n}) = \{X_{0,n} = x_{0,n}\}$, with $x_{0,n} = (x_0, x_1, \dots, x_n)$ ranging over some set $\mathcal{X}_F$ of $n+1$-tuples, then argued that

$$\begin{aligned}&\mathbb{E}_x\left(f(X_{n+1}) - f(X_n) \mid F\right) \\ &\quad = \sum\nolimits_{x_{0,n} \in \mathcal{X}_F} \mathbb{E}_x\left(f(X_{n+1}) - f(X_n) \mid D(x_{0,n}), F\right)\mathbb{P}_x\{D(x_{0,n}) \mid F\}.\end{aligned}$$

Note that

$$\begin{aligned}&\mathbb{E}_x\left(f(X_{n+1}) - f(X_n) \mid D(x_{0,n}), F\right) \\ &\quad = \mathbb{E}_x\left(f(X_{n+1}) - f(X_n) \mid D(x_{0,n})\right) = 0 \qquad \text{for each } x_{0,n} \in \mathcal{X}_F.\end{aligned}$$

*(v) Show that $Z_n$ is also a Markov chain. Write down the state space and transition probabilities for the $Z_n$-chain.*

You could argue formally, using the notation from the previous part, that the transition probabilities were chosen to make

$$\mathbb{P}\{f(X_{n+1}) - f(X_n) = \pm 1 \mid D(x_{0,n})\} = 1/2$$

for each $x_{0,n} = (x_0, \dots, x_n)$ with $x_i \notin A$ for each $i$. An averaging out with weights $\mathbb{P}_x\{D(x_{0,n}) \mid F\}$ would then show that $Z_n = f(X_n)$ is a Markov chain with state space $\mathcal{S}' = \{2, 3, \dots, 2d\}$ and transition probabilities $P(i, i+1) = 1/2 = P(i, i-1)$ for $2 < i < 2d$, with the states 2 and $2d$ absorbing.

*(vi) Define $g(x) = \mathbb{P}_x B$. Show that $g$ is a harmonic function.*

You needed to show that $\mathbb{P}_x g(X_1) = g(x)$ for all $x$ in $\mathcal{S}$. I had intended that you prove this fact by a direct conditioning argument:

$$\mathbb{P}_x g(X_1) = \sum\nolimits_{y \in \mathcal{S}} P(x,y) \mathbb{P}(B \mid X_0 = x, X_1 = y).$$

It might help to think of $B$ as $\cup_{n \in \mathbb{N}} \{X_n = (d,d)\}$, in order to recognize that $\mathbb{P}(B \mid X_0 = x, X_1 = y) = \mathbb{P}_y B = g(y)$. Do we need a separate argument for the cases where $x \in A$? Is there any problem with the case $x = (1,1)$ and $y = (d,d)$?

Many of you actually solved the next part first, using a gambler's ruin argument, then checked manually that the resulting $g$ was harmonic.

*(vii) Find $g(x)$ for each $x$ in $\mathcal{S}$.*

A slick argument could be based on the fact that a harmonic function for the $Z$-chain is uniquely determined by its values $h(2)$ and $h(2d)$. Proof? The harmonic function

$$h(x) = \frac{f(x) - 2}{2(d-1)}$$

has $h(2) = 0 = g(2)$ and $h(2d) = 1 = g(2d)$. It follows that $g(x) = h(x)$ for all $x$ in $\mathcal{S}'$.

[3] *Question 1 on Homework 3 described a modification of the queueing example from Section 2.3 of the Chang notes. I got myself greatly confused over the problem of independence of $X_n$ and $D_1$, $D_2, \ldots,$ $D_n$. The following questions revisit the modified problem, with the aim of proving that*

<1> $$\mathbb{P}_\pi\{X_2 = k, D_1 = \delta_1, D_2 = \delta_2\} = \pi_k \theta(\delta_1)\theta(\delta_2) \qquad \textit{for all } \delta_i \in \{0,1\}, \textit{ all } k \geq 0,$$

*where $\theta(1) = p$ and $\theta(0) = 1 - p$. That is, the aim is to prove independence of $X_2$, $D_1$, and $D_2$. To this end define*

$$G(k, \delta_1, \alpha_1, \delta_2, \alpha_2) = \mathbb{P}_\pi\{X_2 = k, D_1 = \delta_1, A_1 = \alpha_1, D_2 = \delta_2, A_2 = \alpha_2\}.$$

*and let $j = k - \alpha_2 + \delta_2$ and $i = j - \alpha_1 + \delta_1$. Note that $\delta_1$ and $\alpha_1$ are uniquely determined by $i$ and $j$ if $|i - j| = 1$; and $\delta_2$ and $\alpha_2$ are uniquely determined by $j$ and $k$ if $|j - k| = 1$*

*You may assume that the chain has stationary distribution $\pi$ and transition probabilities as shown on the Solutions to Sheet 3.*

The key idea is that the values of $X_n$ and $X_{n-1}$ uniquely determine the values of $A_n$ and $D_n$ when $|X_n - X_{n-1}| = 1$. More precisely,

$$\begin{aligned}\{X_{n-1} = x, X_n = x+1\} \\ &= \{X_{n-1} = x, A_n = 1, D_n = 0\} \\ &= \{X_n = x+1, A_n = 1, D_n = 0\}\end{aligned}$$

and

$$\{X_{n-1} = x, X_n = x-1\}$$
$$= \{X_{n-1} = x, A_n = 0, D_n = 1\}$$
$$= \{X_n = x-1, A_n = 0, D_n = 1\}.$$

That is, when $X_n \neq X_{n-1}$ the information about $A_n$ and $D_n$ is redundant; we have only to retain information about the value of least one of $X_n$ and $X_{n-1}$. Compare with

$$\{X_{n-1} = x, X_n = x\}$$
$$= \{X_{n-1} = x, A_n = 1, D_n = 1\} \cup \{X_{n-1} = x, A_n = 0, D_n = 1\}.$$

*(i) For each $i \geq 0$ define $f_i(\delta) = \mathbb{P}_i\{A_1 = D_1 = \delta\}$ for $\delta \in \{0,1\}$ and $i = 0, 1, 2, \dots$. Write down the expression for $f_i(\delta)$. Hint: You will need to distinguish between the cases $i = 0$ and $i \geq 1$.*

Given the event $X_0 = i$, the random variables $A_1$ and $D_1$ are (conditionally) independent with $\mathbb{P}_i\{A_1 = 1\} = p$ for all $i \geq 0$ and $\mathbb{P}_i\{D_1 = 1\} = a$ for all $i \geq 1$. However $\mathbb{P}_0\{D_1 = 1\} = 0$ because an empty queue contains nobody to be served. Thus

$$f_i(0) = (1-p)(1-a) \qquad \text{and} \qquad f_i(1) = pa \qquad \text{if } i \geq 1$$
$$f_0(0) = 1-p \qquad \text{and} \qquad f_0(1) = 0.$$

*(ii) Explain why*

$$G(k, \delta_1, \alpha_1, \delta_2, \alpha_2)$$
$$= \mathbb{P}_\pi\{X_0 = i, X_1 = j, X_2 = k, D_1 = \delta_1, A_1 = \alpha_1, D_2 = \delta_2, A_2 = \alpha_2\}$$

*for all $k \geq 1$ and all $\delta_i, \alpha_i \in \{0,1\}$.*

Actually, the argument works for all $k \geq 0$.

If we know $X_n$, $A_n$ and $D_n$ then the value of $X_{n-1}$ is uniquely determined. In particular,

$$\{X_2 = k, A_2 = \alpha_2, D_2 = \delta_2\} = \{X_2 = k, A_2 = \alpha_2, D_2 = \delta_2, X_1 = j\}$$

where $j = k - \alpha_2 + \delta_2$, and

$$\{X_1 = j, A_1 = \alpha_1, D_1 = \delta_1\} = \{X_1 = j, A_1 = \alpha_2, D_1 = \delta_1, X_1 = i\}$$

where $i = j - \alpha_1 + \delta_1$. Consequently,

$$\{X_2 = k, D_1 = \delta_1, A_1 = \alpha_1, D_2 = \delta_2, A_2 = \alpha_2\}$$
$$= \{X_2 = k, D_1 = \delta_1, A_1 = \alpha_1, D_2 = \delta_2, A_2 = \alpha_2, X_1 = j, X_0 = i\}.$$

Take $\mathbb{P}_\pi$ probability of both sides.

*(iii) If $\alpha_1 \neq \delta_1$ and $\alpha_2 \neq \delta_2$ show that*

$$\begin{aligned} G(k,&\delta_1, \alpha_1, \delta_2, \alpha_2) \\ &= \mathbb{P}_\pi\{X_0 = k, X_1 = j, X_2 = i\} \\ &= \mathbb{P}_\pi\{X_0 = k, D_1 = \alpha_2, A_1 = \delta_2, D_2 = \alpha_1, A_2 = \delta_1\}. \end{aligned}$$

For this case, $i \neq j$ and $j \neq k$. The values of the $A$'s and $D$'s are uniquely determined by the values of $X_0$, $X_1$, and $X_2$. Thus

$$\begin{aligned} G &= \mathbb{P}_\pi\{X_2 = k, X_1 = j, X_0 = i\} \\ &= \mathbb{P}_\pi\{X_0 = k, X_1 = j, X_2 = i\} \qquad \text{by time reversibility} \\ &= \mathbb{P}_\pi\{X_0 = k, X_1 = j, X_2 = i, A_1 - D_1 = j - k, A_2 - D_2 = i - j\}. \end{aligned}$$

The equality $A_1 - D_1 = j - k = \delta_2 - \alpha_2$ is equivalent to $A_1 = \delta_2$ and $D_1 = \alpha_2$; and $A_2 - D_2 = i - j = \delta_1 - \alpha_1$ is equivalent to $A_2 = \delta_1$ and $D_2 = \alpha_1$. If we make those substitutions then discard the redundant information about $X_1$ and $X_2$ we are left with the asserted expression for $G$.

*(iv) If $\alpha_1 = \delta_1$ and $\delta_2 \neq \alpha_2$ show that*

$$\begin{aligned} G(k,&\delta_1, \alpha_1, \delta_2, \alpha_2) \\ &= \mathbb{P}_\pi\{X_2 = k, X_1 = j, X_0 = j, D_1 = \delta_1 = A_1\} \\ &= \pi_j f_j(\delta_1) P(j,k) \\ &= \mathbb{P}_\pi\{X_0 = k, D_1 = \alpha_2, A_1 = \delta_2, D_2 = \alpha_1, A_2 = \delta_1\}. \end{aligned}$$

For this case, $i = j \neq k$. We need to keep the information about $A_1$ and $D_1$ but the information about $A_2$ and $D_2$ can be replaced by equivalent information about $X_2 - X_1$. Condition.

$$\begin{aligned} G &= \mathbb{P}_\pi\{X_2 = k, X_1 = j, X_0 = j, A_1 = \delta_1, D_1 = \delta_1\} \\ &= \mathbb{P}_\pi\{X_0 = j\} \times \mathbb{P}_\pi\{X_1 = j, A_1 = \delta_1, D_1 = \delta_1 \mid X_0 = j\} \\ &\quad \times \mathbb{P}_\pi\{X_2 = k \mid X_0 = j, X_1 = j, A_1 = \delta_1, D_1 = \delta_1\} \end{aligned}$$

For the last product, the first factor equals $\pi_j$. We can discard the redundant information $X_1 = j$ in the second term to reduce to $f_j(\delta_1)$. The third term equals $P(j,k)$, by the Markov property. Thus $G = \pi_j f_j(\delta_1) P(j,k)$.

Time reversibility gives $\pi_j P(j,k) = \pi_k P(k,j)$. Also, we can reinterpret $f_j(\delta_1)$ as $\mathbb{P}_\pi\{A_2 = D_2 = \delta_1 \mid X_1 = j\}$. With these changes we get

$$\begin{aligned} G &= \pi_k P(k,j) \mathbb{P}_\pi\{A_2 = D_2 = \delta_1 \mid X_1 = j\} \\ &= \mathbb{P}_\pi\{X_0 = k, X_1 = j, A_2 = D_2 = \delta_1\} \end{aligned}$$

We can replace the $X_1 = j$ by the equivalent $A_1 = \delta_2, D_1 = \alpha_2$ because $j - k \neq 0$.

*(v) Similarly, if $\alpha_1 \neq \delta_1$ and $\alpha_2 = \delta_2$, show that*

$$
\begin{aligned}
&G(k,\delta_1, \alpha_1, \delta_2, \alpha_2) \\
&\qquad = \mathbb{P}_\pi\{X_0 = k, D_1 = \alpha_2, A_1 = \delta_2, D_2 = \alpha_1, A_2 = \delta_1\}.
\end{aligned}
$$

For this case, $i \neq j = k$. Argue as in the previous part.

$$
\begin{aligned}
G &= \mathbb{P}_\pi\{X_2 = X_1 = k, X_0 = i, A_2 = D_2 = \delta_2\} \\
&= \pi_i P(i,k) f_k(\delta_2) \\
&= \pi_k \mathbb{P}_\pi\{A_1 = D_1 = \delta_2 \mid X_0 = k\} P(k,i) \qquad \text{time reversibility} \\
&= \mathbb{P}_\pi\{X_0 = k, A_1 = D_1 = \delta_2, X_2 = i\}
\end{aligned}
$$

We can replace the $X_2 = i$ by the equivalent $A_2 = \delta_1, D_2 = \alpha_1$ because $k - i \neq 0$.

*(vi) If $\alpha_1 = \delta_1$ and $\alpha_2 = \delta_2$, show that*

$$
\begin{aligned}
&G(k,\delta_1, \alpha_1, \delta_2, \alpha_2) \\
&\qquad = \mathbb{P}_\pi\{X_2 = k, X_1 = k, X_0 = k, D_1 = \delta_1 = A_1, D_2 = \delta_2 = A_2\} \\
&\qquad = \pi_k f_k(\delta_1) f_k(\delta_2) \\
&\qquad = \mathbb{P}_\pi\{X_0 = k, D_1 = \alpha_2, A_1 = \delta_2, D_2 = \alpha_1, A_2 = \delta_1\}.
\end{aligned}
$$

For this case, $i = j = k$. Condition to factorize the first $\mathbb{P}_\pi$ probability into $\mathbb{P}_\pi\{X_0 = k\}$ times $\mathbb{P}_\pi\{X_1 = k, D_1 = \delta_1 = A_1 \mid X_0 = k\}$ times

$$\mathbb{P}_\pi\{X_2 = k, D_2 = \delta_2 = A_2 \mid X_1 = k, D_1 = \delta_1 = A_1, X_0 = k\}$$

Discard the redundant $X_1 = k$ in the second factor, and the redundant $X_2 = k$ in the third, then invoke the Markov property to reduce the product to $\pi_k f_k(\delta_1) f_k(\delta_2)$. A similar argument reduces

$$\mathbb{P}_\pi\{X_0 = k = X_1 = X_2, D_1 = A_1 = \delta_2, D_2 = A_2 = \delta_1\}$$

to $\pi_k f_k(\delta_2) f_k(\delta_1)$.

*(vii) Complete the proof of <1>.*

For all $k \geq 0$ and all $\alpha_1, \delta_1 \alpha_2, \delta_2$ we now know that

$$
\begin{aligned}
&\mathbb{P}_\pi\{X_2 = k, D_1 = \delta_1, A_1 = \alpha_1, D_2 = \delta_2, A_2 = \alpha_2\} \\
&\qquad = \mathbb{P}_\pi\{X_0 = k, D_1 = \alpha_2, A_1 = \delta_2, D_2 = \alpha_1, A_2 = \delta_1\}.
\end{aligned}
$$

Sum over $\alpha_1$ and $\alpha_2$ to deduce that

$$\mathbb{P}_\pi\{X_2 = k, D_1 = \delta_1, D_2 = \delta_2\} = \mathbb{P}_\pi\{X_0 = k, A_1 = \delta_2, A_2 = \delta_1\}.$$

By independence of the arrival process, the last term factorizes into $\pi_k \theta(\delta_2)\theta(\delta_1)$, which is the desired expression for $\mathbb{P}_\pi\{X_2 = k, D_1 = \delta_1, D_2 = \delta_2\}$.