Corrected proof of stationarity equation for MCMC

As in class, the non-evidence variables are $\mathbf{X} = \{X_1, \ldots, X_n\}$, and the evidence variables $\mathbf{E}$ are set to $\mathbf{e}$. The MCMC algorithm attempts to estimate the conditional distribution of one of the variables, say X_1, given the evidence $\mathbf{e}$, i.e., $\Pr[X_1|\mathbf{e}]$.

We wish to show that the MCMC algorithm takes a random walk whose stationary distribution is given by

$$\pi(\mathbf{x}) = \Pr[\mathbf{X} = \mathbf{x}|\mathbf{e}] = \Pr[\mathbf{x}|\mathbf{e}],$$

meaning that in the long run, the proportion of time steps at which the assignment $\mathbf{x}$ is visited by MCMC is roughly $\pi(\mathbf{x})$. To show this, it suffices to prove the stationarity equation:

$$\pi(\mathbf{x}') = \sum_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \to \mathbf{x}')$$

where $q(\mathbf{x} \to \mathbf{x}')$ is the transition probability of moving from state (assignment) $\mathbf{x}$ to $\mathbf{x}'$. The point of this note is to give a proof of this equation.

We first need to compute this transition probability. This is where I made a mistake in my proof (thanks to Miro for figuring out my bug). If the current assignment is $\mathbf{x}$ and variable X_i is selected, then we change x_i to x_i' with probability

$$\Pr[X_i = x_i'|\mathbf{x}_{-i}, \mathbf{e}]$$

where $\mathbf{x}_{-i}$ is the settings of all the (non-evidence) variables other than X_i. Therefore, in class, I stated that $q(\mathbf{x} \to \mathbf{x}')i$, the transition probability given that variable X_i has been selected, is equal to this probability. However, because no other values of $\mathbf{x}$ are modified, this is true *only if the other values in* $\mathbf{x}'$ *match those in* $\mathbf{x}$; otherwise, the probability is simply zero since there is no chance of making such a transition. In other words,

$$q(\mathbf{x} \to \mathbf{x}'|i) = \begin{cases} \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] & \text{if } \mathbf{x}_{-i} = \mathbf{x}'_{-i} \\ 0 & \text{else.} \end{cases}$$

Since each variable is selected with equal probability, the overall transition probability is

$$q(\mathbf{x} \to \mathbf{x}') = \frac{1}{n} \sum_{i=1}^{n} q(\mathbf{x} \to \mathbf{x}'|i).$$

To prove the stationarity equation, we compute its right hand side:

$$\begin{aligned}
\sum_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \to \mathbf{x}') &= \sum_{\mathbf{x}} \pi(\mathbf{x}) \cdot \frac{1}{n} \sum_{i=1}^{n} q(\mathbf{x} \to \mathbf{x}'|i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \to \mathbf{x}'|i) && \text{by rearranging the sums} \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x}:\mathbf{x}_{-i}=\mathbf{x}'_{-i}} \pi(\mathbf{x}) \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] && \text{plugging in for } q(\mathbf{x} \to \mathbf{x}'|i) \text{ (which is zero when } \mathbf{x}_{-i} \neq \mathbf{x}'_{-i}\text{).}
\end{aligned}$$

As in class, if $\mathbf{x}_{-i} = \mathbf{x}'_{-i}$ then

$$\begin{aligned}
\pi(\mathbf{x}) \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] &= \Pr[\mathbf{x}|\mathbf{e}] \cdot \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] && \text{plugging in for } \pi(\mathbf{x}) \\
&= \Pr[x_i, \mathbf{x}_{-i}|\mathbf{e}] \cdot \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] && \text{decomposing } \mathbf{x} \text{ into } x_i, \mathbf{x}_{-i} \\
&= \Pr[\mathbf{x}_{-i}|\mathbf{e}] \cdot \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}] \cdot \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] && \text{definition of conditional probability} \\
&= \Pr[\mathbf{x}_{-i}|\mathbf{e}] \cdot \Pr[x_i'|\mathbf{x}_{-i}, \mathbf{e}] \cdot \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}] && \text{rearranging factors} \\
&= \Pr[\mathbf{x}'_{-i}|\mathbf{e}] \cdot \Pr[x_i'|\mathbf{x}'_{-i}, \mathbf{e}] \cdot \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}] && \text{since } \mathbf{x}'_{-i} = \mathbf{x}_{-i} \\
&= \Pr[x_i', \mathbf{x}'_{-i}|\mathbf{e}] \cdot \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}] && \text{definition of conditional probability} \\
&= \Pr[\mathbf{x}'|\mathbf{e}] \cdot \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}] && \text{combining } x_i', \mathbf{x}'_{-i} \text{ into } \mathbf{x}' \\
&= \pi(\mathbf{x}') \Pr[x_i|\mathbf{x}_{-i}, \mathbf{e}]. && \text{by definition of } \pi
\end{aligned}$$

So, plugging into the derivation above, we get that the right hand side of the stationarity equation is

$$
\begin{aligned}
\sum_{\mathbf{x}} \pi(\mathbf{x}) q(\mathbf{x} \to \mathbf{x}') &= \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x}:\mathbf{x}_{-i}=\mathbf{x}'_{-i}} \pi(\mathbf{x}') \Pr[x_i | \mathbf{x}_{-i}, \mathbf{e}] \\
&= \pi(\mathbf{x}') \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x}:\mathbf{x}_{-i}=\mathbf{x}'_{-i}} \Pr[x_i | \mathbf{x}_{-i}, \mathbf{e}] && \text{pulling } \pi(\mathbf{x}') \text{ out of the sum} \\
&= \pi(\mathbf{x}') \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathbf{x}:\mathbf{x}_{-i}=\mathbf{x}'_{-i}} \Pr[x_i | \mathbf{x}'_{-i}, \mathbf{e}] && \text{since } \mathbf{x}_{-i} = \mathbf{x}'_{-i} \text{ inside the sum} \\
&= \pi(\mathbf{x}') \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{x_i} \Pr[x_i | \mathbf{x}'_{-i}, \mathbf{e}] && \text{since only } x_i \text{ is changing in the sum, and } \mathbf{x}_{-i} \text{ does not appear inside of it} \\
&= \pi(\mathbf{x}') \cdot \frac{1}{n} \sum_{i=1}^{n} 1 && \text{since the sum of probabilities of all elements of a distribution is 1} \\
&= \pi(\mathbf{x}').
\end{aligned}
$$

This was the desired result showing that the stationarity equation holds.