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

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

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

\[
\pi(x') = \sum_x \pi(x) q(x) P(x \rightarrow x')
\]

where \( q(x \rightarrow x') \) is the transition probability of moving from state (assignment) \( x \) to \( x' \). 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 \( x \) and variable \( X_i \) is selected, then we change \( x_i \) to \( x'_i \) with probability

\[
\Pr[X_i = x'_i | x_{-i}, \mathbf{e}]
\]

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

\[
q(x \rightarrow x') = \begin{cases} 
\Pr[x'_i | x_{-i}, \mathbf{e}] & \text{if } x_{-i} = x'_{-i} \\
0 & \text{else}
\end{cases}
\]

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

\[
q(x \rightarrow x') = \frac{1}{n} \sum_{i=1}^n q(x \rightarrow x')
\]

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

\[
\sum_x \pi(x) q(x \rightarrow x') = \sum_x \pi(x) \cdot \frac{1}{n} \sum_{i=1}^n q(x \rightarrow x')
\]

\[
= \frac{1}{n} \sum_{i=1}^n \sum_x \pi(x) q(x \rightarrow x')
\]

\[
= \frac{1}{n} \sum_{i=1}^n \sum_{x: x_{-i} = x'_{-i}} \pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}]
\]

As in class, if \( x_{-i} = x'_{-i} \) then

\[
\pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}] = \Pr[x_i | \mathbf{e}] \cdot \Pr[x'_i | x_{-i}, \mathbf{e}]
\]

\[
\pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}] = \Pr[x_i, x_{-i} | \mathbf{e}] \cdot \Pr[x'_i | x_{-i}, \mathbf{e}]
\]

\[
\pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}] = \Pr[x_i | x_{-i}] \cdot \Pr[x'_i | x_{-i}, \mathbf{e}]
\]

\[
\pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}] = \Pr[x'_i | x_{-i}] \cdot \Pr[x_i | x_{-i}]
\]

\[
\pi(x) \Pr[x'_i | x_{-i}, \mathbf{e}] = \pi(x) \Pr[x_i | x_{-i}]
\]

by definition of \( \pi \).
So, plugging into the derivation above, we get that the right hand side of the stationarity equation is

\[
\sum_{x} \pi(x)q(x \rightarrow x') = \frac{1}{n} \sum_{i=1}^{n} \sum_{x: x_{-i}=x'_{-i}} \pi(x') \Pr[x_i | x_{-i}, e] \\
= \pi(x') \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{x: x_{-i}=x'_{-i}} \Pr[x_i | x_{-i}, e] \quad \text{pulling } \pi(x') \text{ out of the sum} \\
= \pi(x') \cdot \frac{1}{n} \sum_{i=1}^{n} \sum_{x: x_{-i}=x'_{-i}} \Pr[x_i | x'_{-i}, e] \quad \text{since } x_{-i} = x'_{-i} \text{ inside the sum} \\
= \pi(x') \cdot \frac{1}{n} \sum_{i=1}^{n} \Pr[x_i | x'_{-i}, e] \quad \text{since only } x_i \text{ is changing in the sum,} \\
= \pi(x') \cdot \frac{1}{n} \sum_{i=1}^{n} 1 \quad \text{and } x_{-i} \text{ does not appear inside of it} \\
= \pi(x').
\]

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