Suppose a pair of random variables (X, θ) has joint density $\pi(X, \theta)$ w.r.t. to the product measure $\mathrm{d}X \times \mathrm{d}\theta$ (this can be extended to more complex measures). We wish to draw samples from the marginal density, given by

$$\pi(\theta) = \int \pi(X, \theta)\ \mathrm{d}X. \tag{1}$$

However, the marginal is rarely analytically available due to the integration involved in (1). Nevertheless, we can estimate it unbiasedly using importance sampling

$$\pi(\theta) = \int \pi(X, \theta)\ \mathrm{d}X = \int \frac{\pi(X, \theta)}{q_\theta(X)}\ q_\theta(X)\mathrm{d}X, \tag{2}$$

where $q_\theta(X)$ is an appropriate importance sampling density. An ubiased estimate is then given by

$$\begin{aligned}\hat{\pi}_N(\theta) &:= \frac{1}{N}\sum_{i=1}^{N}\frac{\pi(X_i, \theta)}{q_\theta(X_i)} \\ &= \frac{1}{N\prod_{j=1}^{N} q_\theta(X_j)}\sum_{i=1}^{N}\left(\pi(X_i, \theta)\prod_{j\neq i}^{N} q_\theta(X_j)\right)\end{aligned} \tag{3}$$

where $X_i \sim q_\theta(X)$. For notational simplicity we let $\mathbf{X} := (X_1, \ldots, X_N)$, denote its joint density by $q_\theta^N(\mathbf{X}) = \prod_{j=1}^{N} q_\theta(X_j)$ and let

$$\hat{\pi}_N(\mathbf{X}, \theta) := \frac{1}{N}\sum_{i=1}^{N}\left(\pi(X_i, \theta)\prod_{j\neq i}^{N} q_\theta(X_j)\right) \tag{4}$$

Therefore, (3) becomes

$$\hat{\pi}_N(\theta) = \frac{\hat{\pi}_N(\mathbf{X}, \theta)}{q_\theta^N(\mathbf{X})} \tag{5}$$

If the target density $\pi(\theta)$ was known analytically then we could apply a M-H algorithm to draw samples as in **Algorithm 1** (see below). In the long run (and under mild assumptions) the algorithm will be drawing samples from the target density.

Since the target density is unknown, an idea is to replace it with the unbiased estimate in (5). This is essentially the idea of the pseudo-marginal approach.

Algorithm 1 The idealized M-H.

1: Propose a value $\theta^* \sim q(\theta, .)$
2: Calculate the acceptance probability

$$p := 1 \wedge \frac{\pi(\theta^*)q(\theta^*, \theta)}{\pi(\theta)q(\theta, \theta^*)} \tag{6}$$

3: With probability p set $\theta = \theta^*$, otherwise set $\theta = \theta$.
4: Go to 1.

If $\mathbf{X}^* \sim q^N_{\theta^*}(.)$ and $\mathbf{X} \sim q^N_{\theta}(.)$, then replacing the unknown marginal density in the numerator and denumerator of (6) we get

$$\begin{aligned} & 1 \wedge \frac{\hat{\pi}_N(\theta^*)q(\theta^*, \theta)}{\hat{\pi}_N(\theta)q(\theta, \theta^*)} \\ & 1 \wedge \frac{\left[\hat{\pi}_N(\mathbf{X}^*, \theta^*)/q^N_{\theta^*}(\mathbf{X}^*)\right] q(\theta^*, \theta)}{\left[\hat{\pi}_N(\mathbf{X}, \theta)/q^N_{\theta}(\mathbf{X})\right] q(\theta, \theta^*)} \\ & 1 \wedge \frac{\hat{\pi}_N(\mathbf{X}^*, \theta^*)q^N_{\theta}(\mathbf{X})q(\theta^*, \theta)}{\hat{\pi}_N(\mathbf{X}, \theta)q^N_{\theta^*}(\mathbf{X}^*)q(\theta, \theta^*)} \end{aligned} \tag{7}$$

Assume for a moment that $\hat{\pi}_N(\mathbf{X}, \theta)$ is a well defined joint density of the pair of random variables $\mathbf{Z} := (\mathbf{X}, \theta)$. We then observe that the above acceptance probability resembles a M-H for $\mathbf{Z}$ where at each step of the algorithm proposals $\mathbf{Z}^*$ are made from $q(\theta, \theta^*)q^N_{\theta^*}(\mathbf{X}^*)$. A pseudo-algorithm can be found in **Algorithm 2**.

Algorithm 2 The pseudo-marginal approach.

1: Propose a value $\theta^* \sim q(\theta, .)$ and $\mathbf{X}^* \sim q^N_{\theta^*}(.)$.
2: Calculate the acceptance probability

$$\hat{p} := 1 \wedge \frac{\hat{\pi}_N(\mathbf{X}^*, \theta^*)q^N_{\theta}(\mathbf{X})q(\theta^*, \theta)}{\hat{\pi}_N(\mathbf{X}, \theta)q^N_{\theta^*}(\mathbf{X}^*)q(\theta, \theta^*)} \tag{8}$$

3: With probability $\hat{p}$ set $\{\theta = \theta^*, \mathbf{X} = \mathbf{X}^*\}$, otherwise set $\{\theta = \theta, \mathbf{X} = \mathbf{X}\}$.
4: Go to 1.

Assuming that $\hat{\pi}_N(\mathbf{X},\theta)$ is a proper joint density then the above algorithm (in the long run) will be drawing samples from this joint density. Therefore, all we need to show is that $\hat{\pi}_N(\mathbf{X},\theta)$ is indeed a density. It turns out that $\hat{\pi}_N(\mathbf{X},\theta)$ is a joint density with respect to the product measure $\mathrm{d}X_1 \times \ldots \mathrm{d}X_N \times \mathrm{d}\theta$. For simplicity, we show it only for $N = 2$, the proof is easily extended for any value N.

$$
\begin{aligned}
&\int \hat{\pi}_N(\mathbf{X},\theta)\mathrm{d}X_1\mathrm{d}X_2 = \frac{1}{2}\int \left\{\pi(X_1,\theta)q_\theta(X_2) + \pi(X_2,\theta)q_\theta(X_1)\right\}\mathrm{d}X_1\mathrm{d}X_2 \\
&= \frac{1}{2}\int \left\{q_\theta(X_2)\int \pi(X_1,\theta)\mathrm{d}X_1 + \pi(X_2,\theta)\int q_\theta(X_1)\mathrm{d}X_1\right\}\mathrm{d}X_2 \\
&= \frac{1}{2}\int \left[q_\theta(X_2)\pi(\theta) + \pi(X_2,\theta)\right]\mathrm{d}X_2 \\
&= \frac{1}{2}\left[\pi(\theta)\int q_\theta(X_2)\mathrm{d}X_2 + \int \pi(X_2,\theta)\mathrm{d}X_2\right] \\
&= \frac{1}{2}\left[\pi(\theta) + \pi(\theta)\right] \\
&= \pi(\theta)
\end{aligned}
$$

Therefore, not only does $\hat{\pi}_N(\mathbf{X},\theta)$ defines a proper density, but also its marginal coincides with the target density $\pi(\theta)$. Therefore, the pseudo-marginal approach (in the long run) draws samples from the correct target density.