Supplementary Material

Proof of Theorem 1: Follow the proof of Theorem 6 in (Maurer & Pontil, 2009) with the deterministic function $f_\pi(x, a)$ given by (3) and the relations between f_π and v^π_Φ given by (4).

Proof Sketch of Theorem 2: Write $\hat{v}^\pi_\Phi$ as the estimated value from (2) given the dataset $\mathcal{D}$, policy $\pi \in \Pi$, and context distribution $\mu(x)$. Given the predicted reward $\bar{r}(a, x)$ for $(a, x) \notin \mathcal{D}$, the estimated value of a policy π can be computed using:

$$\hat{v}^\pi_\Phi = \sum_{a,x} \mu(x)\pi(a, x)\big(\mathbf{1}\{n(a, x) = 0\}\bar{r}(a, x) + \mathbf{1}\{n(a, x) \neq 0\}\hat{r}(a, x)\big)$$

with $n(a, x) = \sum_{t=1}^{T} \mathbf{1}\{a_t = a, x_t = x\}$. Take the expected value of $\hat{v}^\pi_\Phi$ and utilize the law of iterative expectations to obtain

$$\mathbb{E}[\hat{v}^\pi_\Phi] = \sum_{a,x} \mu(x)\pi(a, x)\big(\mathbb{P}(n(a, x) = 0)\bar{r}(a, x) + \mathbb{P}(n(a, x) \neq 0)\hat{r}(a, x)\big). \tag{1}$$

Given the policy $\pi(\cdot|x)$ and expected rewards $\hat{r}(a, x)$ the actual value of a policy π is given by:

$$v^\pi_\Phi = \sum_{a,x} \mu(x)\pi(a, x)\big(\mathbb{P}(n(a, x) = 0) + \mathbb{P}(n(a, x) \neq 0)\big)r(a, x). \tag{2}$$

Subtracting (1) from (2), we obtain the result.

Proof Sketch of Theorem 3: Use the law of total variance $\mathbb{V}[\hat{v}^\pi_\Phi] = \mathbb{E}[\mathbb{V}[\hat{v}^\pi_\Phi|(a, x) \in \mathcal{D}]] + \mathbb{V}[\mathbb{E}[\hat{v}^\pi_\Phi|(a, x) \in \mathcal{D}]]$ and apply the same methods as the proof for Theorem 2.