# Supplement Material to "Efficient Bayes Risk Estimation for Cost-Sensitive Classification", AISTATS 2019

Daniel Andrade and Okajima Yuzuru
Security Research Laboratories, NEC

## 1 Proofs

**Theorem 1.** *The procedure* $\pi^*$ *is a Bayes procedure, that means for any other decision procedure* $\pi$ *we have*

$$\mathbb{E}_{\mathbf{x},y}[l((\mathbf{x},y),\pi^*)] \leq \mathbb{E}_{\mathbf{x},y}[l((\mathbf{x},y),\pi)]\,. \tag{1}$$

*Proof.* Let $S \subseteq V$ be the set of already observed covariates, then the expected remaining costs for a decision procedure $\pi$ is given by

$$\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi,S)|\mathbf{x}_S]\,.$$

We will prove by induction that for any $S \subseteq V$, and any decision procedure $\pi$, we have

$$\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] \leq \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi,S)|\mathbf{x}_S]\,.$$

The claim then follows by setting $S := \emptyset$.

**Base case: $S = V$.** We have

$$\pi^*(\mathbf{x}_S) = \arg\min_{i \in L} \mathbb{E}_y[c_{y,i}|\mathbf{x}_S]\,.$$

Therefore $\pi^*(\mathbf{x}_S)$ is a Bayes procedure, and as a consequence

$$\mathbb{E}_y[c_{y,\pi^*(\mathbf{x}_S)}|\mathbf{x}_S] \leq \mathbb{E}_y[c_{y,\pi(\mathbf{x}_S)}|\mathbf{x}_S]\,.$$

And therefore

$$\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] \leq \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi,S)|\mathbf{x}_S]\,.$$

(Since $S = V$, and we have

$$\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] = \mathbb{E}_y[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] = \mathbb{E}_y[c_{y,\pi^*(S)}|\mathbf{x}_S]\,,$$

and the same analogously for $\pi$.)

**Induction step:** $S \subset V$. Assume that for all $S \cup \{i\}$, where $i \in V \setminus S$, the induction assumptions holds, that is

$$\mathbb{E}_{\mathbf{x}_{V\setminus(S\cup\{i\})},y}[l((\mathbf{x},y),\pi^*,S\cup\{i\})|\mathbf{x}_{S\cup\{i\}}] \leq \mathbb{E}_{\mathbf{x}_{V\setminus(S\cup\{i\})},y}[l((\mathbf{x},y),\pi,S\cup\{i\})|\mathbf{x}_{S\cup\{i\}}]\,.$$

Let $\hat{\pi}$ denote a Bayes procedure. Using the structure of the loss function as defined in the main article, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\hat{\pi},S)|\mathbf{x}_S] &= \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[\begin{cases} c_{y,\hat{\pi}(\mathbf{x})} & \text{if } \hat{\pi}(\mathbf{x}) \in L\,, \\ c_{\hat{\pi}(\mathbf{x})} + l((\mathbf{x},y),\hat{\pi},S\cup\{\hat{\pi}(\mathbf{x})\}) & \text{else.} \end{cases} |\mathbf{x}_S] \\
&\geq \min_{i\in L\cup(V\setminus S)} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[\begin{cases} c_{y,i} & \text{if } i \in L\,, \\ c_i + l((\mathbf{x},y),\hat{\pi},S\cup\{i\}) & \text{else.} \end{cases} |\mathbf{x}_S] \\
&= \min_{i\in L\cup(V\setminus S)} \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\hat{\pi},S\cup\{i\})|\mathbf{x}_S] & \text{else.} \end{cases} \\
&= \min_{i\in L\cup(V\setminus S)} \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{x_i}\left[\mathbb{E}_{\mathbf{x}_{V\setminus(S\cup\{i\})},y}\left[l((\mathbf{x},y),\hat{\pi},S\cup\{i\})|\mathbf{x}_{S\cup\{i\}}\right]|\mathbf{x}_S\right] & \text{else.} \end{cases} \\
&\overset{(1)}{\geq} \min_{i\in L\cup(V\setminus S)} \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{x_i}\left[\mathbb{E}_{\mathbf{x}_{V\setminus(S\cup\{i\})},y}\left[l((\mathbf{x},y),\pi^*,S\cup\{i\})|\mathbf{x}_{S\cup\{i\}}\right]|\mathbf{x}_S\right] & \text{else.} \end{cases} \\
&= \min_{i\in L\cup(V\setminus S)} \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}\left[l((\mathbf{x},y),\pi^*,S\cup\{i\})|\mathbf{x}_S\right] & \text{else.} \end{cases} \\
&= \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S]\,,
\end{aligned}$$

where in the line marked by (1) we used the induction assumption. The last line follows from Lemma 1. Since $\hat{\pi}$ is a Bayes procedure, we must have equality in the second and fifth line. Therefore $\pi^*$ is also a Bayes procedure. □

**Lemma 1.**

$$\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] = \min_{i\in L\cup(V\setminus S)} \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,i}|\mathbf{x}_S] & \textit{if } i \in L\,, \\ c_i + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}\left[l((\mathbf{x},y),\pi^*,S\cup\{i\})|\mathbf{x}_S\right] & \textit{else.} \end{cases}$$

*Proof.*

$$\begin{aligned}
\mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S)|\mathbf{x}_S] &= \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[\begin{cases} c_{y,\pi^*(\mathbf{x}_S)} & \text{if } \pi^*(\mathbf{x}_S) \in L\,, \\ c_{\pi^*(\mathbf{x}_S)} + l((\mathbf{x},y),\pi^*,S\cup\{\pi^*(\mathbf{x}_S)\}) & \text{else.} \end{cases} |\mathbf{x}_S] \\
&= \begin{cases} \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[c_{y,\pi^*(\mathbf{x}_S)}|\mathbf{x}_S] & \text{if } \pi^*(\mathbf{x}_S) \in L\,, \\ c_{\pi^*(\mathbf{x}_S)} + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S\cup\{\pi^*(\mathbf{x}_S)\})|\mathbf{x}_S] & \text{else.} \end{cases} \\
&= \begin{cases} \mathbb{E}_{y}[c_{y,\pi^*(\mathbf{x}_S)}|\mathbf{x}_S] & \text{if } \pi^*(\mathbf{x}_S) \in L\,, \\ c_{\pi^*(\mathbf{x}_S)} + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S\cup\{\pi^*(\mathbf{x}_S)\})|\mathbf{x}_S] & \text{else.} \end{cases}
\end{aligned}$$

**1. Case:** $\pi^*(\mathbf{x}_S) \in L$**.** Then because of the definition of $\pi^*$, we have

$$\mathbb{E}_y[c_{y,\pi^*(\mathbf{x}_S)}|\mathbf{x}_S] = \min_{i \in L \cup (V \setminus S)} \begin{cases} \mathbb{E}_y[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S \cup \{i\})|\mathbf{x}_S] & \text{else.} \end{cases}$$

**2. Case:** $\pi^*(\mathbf{x}_S) \notin L$**.** Then because of the definition of $\pi^*$, we have

$$\begin{aligned} & c_{\pi^*(\mathbf{x}_S)} + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S \cup \{\pi^*(\mathbf{x}_S)\})|\mathbf{x}_S] \\ &= \min_{i \in L \cup (V \setminus S)} \begin{cases} \mathbb{E}_y[c_{y,i}|\mathbf{x}_S] & \text{if } i \in L\,, \\ c_i + \mathbb{E}_{\mathbf{x}_{V\setminus S},y}[l((\mathbf{x},y),\pi^*,S \cup \{i\})|\mathbf{x}_S] & \text{else.} \end{cases} \end{aligned}$$

□